**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.

***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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Small for gestational age (n = 105) | Normal for gestational age (n = 650) | Overall (n=755) | Number missing |
| Height (cm)1 | 155 (5.87, 142-172) | 157 (6.54, 106-176) | 157 (6.50, 106-176) | 6 |
| Age (yrs)1 | 23.9 (4.90, 16-35) | 24.9 (5.45, 14-43) | 24.8 (5.39, 14-43) | 0 |
| Parity1 | 0.90 (1.11, 0-6) | 1.13 (1.23, 0-6) | 1.10 (1.21, 0-6) | 0 |
| Smoker2 | 43% | 29% | 31% | 4 |
| Birth Weight (kg)1 | 2.23 (0.41, 1.04-3.78) | 3.25 (0.40, 2.51-4.73) | 3.11 (0.53, 1.04-4.73) | 4 |
| Child is Female2 | 58% | 48% | 49% | 4 |
| Gestational Age (weeks)1 | 37.9 (2.20, 30-42) | 39.4 (1.25, 38-44) | 39.2 (1.50, 30-44) | 5 |

1 Descriptive statistics provided are: mean (standard deviation, min-max)

2 Descriptive statistics provided are: frequency

Methods: For continuous variables (height, age, birth weight, gestational age) and for the ordered categorical variable parity the mean, standard deviation, min, and max were calculated for the groups of participants who had children that were small for gestational age, normal for gestational age, and overall. For binary variables (smoking status and sex of child) the frequency was provided for each of the previously mentioned groups.

Results: It can be seen from the table that women who gave birth to children who were small for gestational age had a lower mean height, lower mean age, lower mean parity, were more likely to smoke, had children with a lower birth weight, were more likely to give birth to a female child, and had a shorter gestational age than those women who gave birth to children who were normal for gestational age. It is possible that there is bias present due to the participants with missing data.

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: The odds of delivering a child who is small for gestational age (SGA) based on smoking status was examined using logistic regression. 4 participants were excluded from the analysis since they were missing data. 95% Confidence intervals were calculated using robust standard errors.

Results: Of the 231 smokers the odds of having a SGA infant was 0.25 and of the 520 nonsmokers the odds of having a SGA infant was 0.13, giving an odds ratio of 1.89. 95% confidence intervals suggest that it would not be unusual if the true odds ratio was between 1.24 and 2.89. A two-sided p value (p=0.003) shows that this observation is statistically significant at the 0.05 level. Therefore, we can reject the null hypothesis that there is no association between smoking and SGA.

* 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.

Table of Model Parameters:

|  |  |  |
| --- | --- | --- |
|  | Estimate | exp(Est) |
| Intercept | -2.06 | 0.13 |
| Slope | 0.64 | 1.89 |

The odds of delivering a SGA infant for smokers is 0.24. This is the odds ratio, the exponentiated estimate for the intercept, multiplied by the odds of delivering a SGA infant for nonsmokers, the exponentiated estimate for the slope (1.89 \* 0.13 = 0.24)1.

The probability of delivering a SGA infant for smokers is 19.2%. This can be calculated by knowing the number of smokers (n = 231). The odds of delivering a SGA infant for smokers (0.24) plus 1 all multiplied by some factor x equals the number of smokers ( x\*(0.24+1) = 231). x\*0.25 = the number of smokers who delivered a SGA infant (x =186; 186\*0.24 = 45)1. x = the number of smokers who did not deliver a SGA infant (x = 186). We can then divide the number of smokers who delivered a SGA infant by the total number of smokers to find the probability of delivering a SGA infant for smokers (45/231 = 19.2%).

The odds of delivering a SGA infant for nonsmokers is 0.13. This is exactly the exponentiated estimate of the intercept, which is the odds of having a SGA infant if the participant is a nonsmoker.

The probability of delivering a SGA infant for nonsmokers is 11.3%. This can be calculated by knowing the number of nonsmokers (n = 520). The odds of delivering a SGA infant for nonsmokers (0.13) plus 1 all multiplied by some factor x equals the number of smokers ( x\*(0.13+1) = 520). x\*0.13 = the number of smokers who delivered a SGA infant (x = 461; 461\*0.13 = 60)1. x = the number of smokers who did not deliver a SGA infant (x = 464). We can then divide the number of smokers who delivered a SGA infant by the total number of smokers to find the probability of delivering a SGA infant for smokers (59/520 = 11.3%).

All of these estimates can be directly related to the descriptive statistics in the table in question 1. Using the frequency of smokers in the SGA and non-SGA group a 2x2 table can be created and the odds, probabilities, and odds ratio can be calculated. They are all nearly identical to the estimated values if all calculations are carried out without rounding till the final value.

104 \* 0.43 = 45 (Note: one missing value for smoking in the SGA group)

104 – 45 = 59

647 \* 0.29 = 186 (Note: three missing values for smoking in the non-SGA group. Also, full value used for frequency, 0.2875)

647 – 186 = 461

|  |  |  |  |
| --- | --- | --- | --- |
|  | **SGA** | **non-SGA** | **Row Totals** |
| **Smokers** | 45 | 186 | 231 |
| **Nonsmokers** | 59 | 461 | 520 |
| **Column Totals** | 104 | 647 |  |

Odds (smoker) = 45/186 = 0.24

Probability (smoker) = 45/231 = 19.2%

Odds (nonsmoker) = 59/461 = 0.13

Probability (nonsmoker) = 59/520 = 11.3%

OR = 0.24/0.13 = 1.89

1 Note that the calculations as shown are only approximate, however when carried out with proper precision they produce the correct numbers.

* 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.

If the value for smoking status were inverted then the exponentiated estimation for the intercept would be the odds of delivering a SGA infant for a smoker. The exponentiated estimation for the slope would be the odds ratio of SGA delivery for nonsmokers over smokers and would be less than 1.

* + 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.

If the SGA indicator variable was inverted then the exponentiated estimation for the intercept would be the odds of delivering a non-SGA infant for a nonsmoker. The exponentiated estimation for the slope would be the odds ratio of non-SGA delivery for smokers over nonsmokers and would be less than 1.

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

If the SGA indicator variable was inverted and the smoking status was inverted then the exponentiated estimation for the intercept would be the odds of delivering a non-SGA infant for a smoker. The exponentiated estimation for the slope would be the odds ratio of non-SGA delivery for nonsmokers over smokers and would be greater than 1.

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

Methods: The difference in probabilities of delivering a child who is small for gestational age (SGA) based on smoking status was examined using linear regression. 4 participants were excluded from the analysis since they were missing data. 95% Confidence intervals were calculated using robust standard errors.

Results: The probability of having a SGA infant for smokers (n = 231) was 19.4% and the probability of having a SGA infant for nonsmokers (n = 520) was 11.3%, leading to a difference in probabilities of 8.1%. 95% confidence intervals suggest that it would not be unusual if the true difference in probabilities was between 2.3% and 14%. A two-sided p value (p=0.006) shows that this observation is statistically significant at the 0.05 level. Therefore, we can reject the null hypothesis that there is no association between smoking and SGA.

* 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.

Table of Model Parameters:

|  |  |
| --- | --- |
|  | Estimate |
| Intercept | 0.113 |
| Slope | 0.081 |

The odds of delivering a SGA infant for smokers is 0.24. The total number of smokers (231) must be known then the number of nonsmokers who delivered a SGA infant can be found (231 \* 0.194 = 45)1 and the number of nonsmokers who did not deliver a SGA infant can be found (231 – 45 = 186). From this the odds can be calculated as 45/186 = 0.24.

The probability of delivering a SGA infant for smokers is 19.4%. This can be found by adding the estimation for the slope to the estimation for the intercept (11.3% + 8.1% = 19.4%). This is slightly different than that predicted by the logistic regression and the value calculated from the descriptive statistics in the 2x2 table.

The odds of delivering a SGA infant for nonsmokers is 0.13. The total number of nonsmokers (520) must be known then the number of nonsmokers who delivered a SGA infant can be found (520 \* 0.113 = 59) and the number of nonsmokers who did not deliver a SGA infant can be found (520 – 59 = 461). From this the odds can be calculated as 59/461 = 0.13.

The probability of delivering a SGA infant for nonsmokers is 11.3%. This is the estimation of the intercept (11.3%).

All of these estimates can be directly related to the descriptive statistics in the table in question 1. Using the frequency of smokers in the SGA and non-SGA group a 2x2 table can be created and the odds, probabilities, and odds ratio can be calculated. They are all nearly identical to the estimated values if all calculations are carried out without rounding till the final value.

1 Note that the calculations as shown are only approximate, however when carried out with proper precision they produce the correct numbers.

c.i) If the value for smoking status were inverted then the estimation for the intercept would be the probability of delivering a SGA infant for a smoker. The estimation for the slope would be the difference in probabilities of SGA delivery for nonsmokers over smokers and would be less than 1.

c.ii) If the SGA indicator variable was inverted then the estimation for the intercept would be the probability of delivering a non-SGA infant for a nonsmoker. The estimation for the slope would be the difference in probabilities of non-SGA delivery for smokers over nonsmokers and would be less than 1.

c.iii) If the SGA indicator variable was inverted and the smoking status was inverted then the estimation for the intercept would be the probability of delivering a non-SGA infant for a smoker. The estimation for the slope would be the difference in probabilities of non-SGA delivery for nonsmokers over smokers and would be greater than 1.

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

Methods: The ratio of the probabilities of delivering a child who is small for gestational age (SGA) based on smoking status was examined using Poisson regression. 4 participants were excluded from the analysis since they were missing data. 95% Confidence intervals were calculated using robust standard errors.

Results: The rate of delivering a SGA infant was found to be 1.72 times greater in smokers compared to nonsmokers. 95% confidence intervals suggest that it would not be unusual if the true ratio in probabilities was between 1.20 and 2.45. A two-sided p value (p=0.003) shows that this observation is statistically significant at the 0.05 level. Therefore, we can reject the null hypothesis that there is no association between smoking and SGA.

* 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.

Table of Model Parameters:

|  |  |  |
| --- | --- | --- |
|  | Estimate | exp(Est) |
| Intercept | -2.18 | 0.11 |
| Slope | 0.54 | 1.72 |

The original 2x2 table can be recovered from the Poisson regression data if the total number of smokers (n=231) and nonsmokers (n=520) are known. The number of nonsmokers who delivered a SGA infant is the number of nonsmokers times the exponentiated estimate of the intercept (520 \* 0.11 = 59). The number of nonsmokers who delivered a non-SGA infant is the total number of nonsmokers minus the number of nonsmokers who delivered a SGA infant (520 – 59 = 461). The number of smokers who delivered a SGA infant is the number of smokers times the exponentiated estimate of the intercept times the exponentiated estimate of the slope (231 \* 0.11 \* 1.79 = 45; also note that 0.11 \* 1.79 = 0.19, the probability of having a SGA infant for a smoker). The number of smokers who delivered a non-SGA infant is the total number of smokers minus the number of smokers who delivered a SGA infant (231 – 45 = 186).

The odds of delivering a SGA infant for smokers is 0.24. From the above information the odds can be calculated as 45/186 = 0.24.

The probability of delivering a SGA infant for smokers is 18.9%. This can be found by multiplying the exponentiated estimation for the slope to the exponentiated estimation for the intercept (1.79 \* 0.11 = 0.189). This is slightly different than that predicted by the logistic regression, linear regression and the value calculated from the descriptive statistics in the 2x2 table. This difference may be due to truncating the value during calculations.1

The odds of delivering a SGA infant for nonsmokers is 0.13. From the above information the odds can be calculated as 59/461 = 0.13.

The probability of delivering a SGA infant for nonsmokers is 11%. This is the exponentiated estimation of the intercept (11%).

All of these estimates can be directly related to the descriptive statistics in the table in question 1. Using the frequency of smokers in the SGA and non-SGA group a 2x2 table can be created and the odds, probabilities, and odds ratio can be calculated. They are all nearly identical to the estimated values if all calculations are carried out without rounding till the final value.

1 Note that the calculations as shown are only approximate, however when carried out with proper precision they produce the correct numbers.

c.i) If the value for smoking status were inverted then the exponentiated estimation for the intercept would be the probability of delivering a SGA infant for a smoker. The exponentiated estimation for the slope would be the ratio of the probabilities of SGA delivery for nonsmokers over smokers and would be less than 1.

c.ii) If the SGA indicator variable was inverted then the exponentiated estimation for the intercept would be the probability of delivering a non-SGA infant for a nonsmoker. The exponentiated estimation for the slope would be the ratio of the probabilities of non-SGA delivery for smokers over nonsmokers and would be less than 1.

c.iii) If the SGA indicator variable was inverted and the smoking status was inverted then the exponentiated estimation for the intercept would be the probability of delivering a non-SGA infant for a smoker. The exponentiated estimation for the slope would be the ratio of the probabilities of non-SGA delivery for nonsmokers over smokers and would be greater than 1.

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?

The linear regression is predicting the difference in proportions of SGA delivery between smokers and nonsmokers, the logistic regression is predicting the multiplicative increase in SGA (odds ratio) between smokers and nonsmokers, the Poisson regression is predicting the multiplicative increase in the rate of SGA (ratio of proportions) between smokers and nonsmokers. The simple two sample comparison, if performed with a two-sample t-test allowing for unequal variances, would be the same as the linear regression with robust standard errors. If a simple two sample comparison was performed using the chi-square test then this would be similar to the logistic regression (score statistic). And if the simple two sample comparison was a two sample test of Poisson rates then it would similar to the Poisson regression.

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: The difference in probabilities of delivering a child who is small for gestational age (SGA) based on maternal age was examined using linear regression. 95% Confidence intervals were calculated using robust standard errors.

Results: Linear regression was performed and it showed that every year increase in maternal age leads to a 0.45% decrease in the probability of delivery of a SGA infant. 95% confidence intervals suggest that it would not be unusual if the true decrease in probability was between 0.029% and 0.87%. A two-sided p value (p=0.04) shows that this observation is statistically significant at the 0.05 level. Therefore, we can reject the null hypothesis that there is no association between maternal age and SGA.

Table for question 7:

|  |  |
| --- | --- |
|  | Estimate |
| Intercept | 0.25 |
| Slope | -0.0045 |

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

Methods: The ratio of the probabilities of delivering a child who is small for gestational age (SGA) based on maternal age was examined using Poisson regression. 95% Confidence intervals were calculated using robust standard errors.

Results: Poisson regression was performed and it was found that every 1 year increase in age leads to a 3.4% decrease in the rate of SGA delivery in the older group compared to the younger group.1 95% confidence intervals suggest that it would not be unusual if the decrease in the SGA delivery rate per 1 year increase in age was between 0.05% and 6.6%. A two-sided p value (p=0.05) shows that this observation is not statistically significant at the 0.05 level. Therefore, we cannot reject the null hypothesis that there is no association between maternal age and SGA.

1 The exponentiated slope estimate was 0.966, meaning a 0.966-fold increase in SGA delivery rate per 1 year difference in age. This can also be written as 100 \* (1-0.966) = 3.4% decrease in SGA delivery rate per 1 year increase in age.

Table for question 7:

|  |  |
| --- | --- |
|  | Estimate |
| Intercept | -1.14 |
| Slope | -0.034 |

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

Methods: The ratio of the odds of delivering a child who is small for gestational age (SGA) based on maternal age was examined using logistic regression. 95% Confidence intervals were calculated using robust standard errors.

Results: Through a logistic regression analysis it was found that every 1 year increase in maternal age is associated with a 4% decrease in the odds of delivery of a SGA infant in the older group compared to the younger group. 95% confidence intervals suggest that it would not be unusual if the true decrease in the odds of delivery of a SGA infant was between 0% and 8%. A two-sided p value (p=0.05) shows that this observation is not statistically significant at the 0.05 level. Therefore, we cannot reject the null hypothesis that there is no association between maternal age and SGA.

Table for question 7:

|  |  |
| --- | --- |
|  | Estimate |
| Intercept | -0.85 |
| Slope | -0.040 |

* 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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Linear | Logistic | Poisson | Sample Proportion |
| Estimate | 16% | 19.2% | 16.2% | 7.5% |

Linear estimate: SGA = -.0045\*age + 0.25

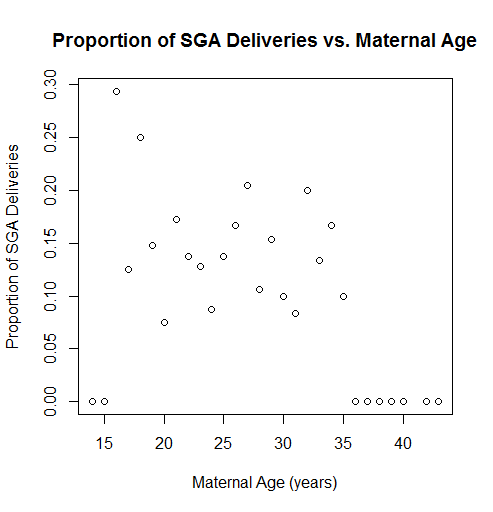
Logistic estimate: log SGA = -0.040\*age – 0.85; SGA = exp(-0.040\*age – 0.85)

Poisson estimate: Event rate of SGA = exp(-.034\*age – 1.14)

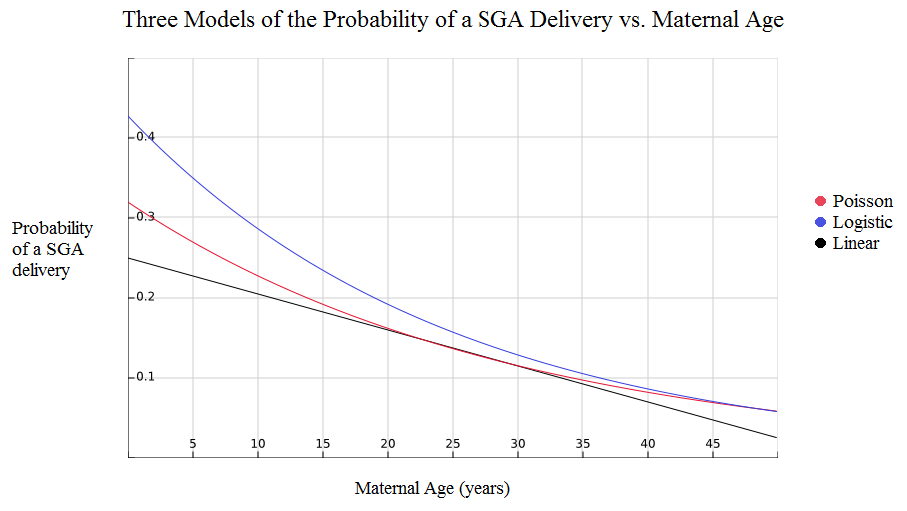
Sample proportion: SGA deliveries for 20 year olds / total deliveries for 20 year olds = 3/40

The regressions differ substantially from the sample proportion since there are only 40 20-year-olds in the dataset. The regression analyses use data from other age groups to estimate the probability of a SGA delivery and therefore have, in essence, a “larger sample size”.

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 from 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: This can be obtained in Stata using the command egen *varname*= mean(sga), by(age).



* 1. Estimated probabilities for each age in the data as derived from each of the regression analyses. In Stata, this can be obtained using the simple “post-estimation” command: predict *varname.* (But use a different variable name for each fitted value.)
     1. After performing a linear regression, the default action of the “predict” function is to create a variable that contains the estimated “linear predictor”, which corresponds to the regression based estimate of the mean. With a binary response variable, the mean response is the proportion.
     2. After performing a Poisson regression, the default action of the “predict” function is to create a variable that contains the exponentiated estimated “linear predictor”, which corresponds to the regression based estimate of the mean. With a binary response variable, the mean response is the proportion. (The linear predictor in Poisson regression corresponds to the log “rate”, because Poisson regression uses a log link function.
     3. In logistic regression, the estimated “linear predictor” corresponds to the log odds. Exponentiating that would correspond to the odds. By default, Stata figures that you would really rather have the estimated probability, which is computed as prob = odds / (1 + odds). So, after performing a logistic regression, the default action of the “predict” function is to create a variable that contains the regression based estimate of the mean.



One of the things that is immediately apparent when looking at the above graph are the y-intercepts and x-intercepts (not shown). Each of the regressions has a unique y-intercept, predicting a different probability for a SGA delivery given a maternal age of 0 years. Of course it doesn't make sense to talk about pregnancy and delivery till maternal age is >10 years or so (14 years was the youngest delivery in this dataset). The x-intercept for the linear regression is around 55 years , however the logistic and Poisson regressions do not every intersect the x-axis since they are decaying exponentials with a horizontal asymptote at y=0. This means that they still predict some non-zero probability of SGA delivery for all maternal ages. Of course it doesn't make sense to talk about pregnancy and delivery for maternal ages over about 45 years or so (43 years was the oldest delivery in this dataset). The logistic regression follows the linear regression closely for maternal ages between approximately 17 years and 37 years (more or less the normal span of pregnancy and delivery). However, as we saw in question 6 part d, the logistic model predicts a higher probability of SGA delivery than the Poisson and linear models. This holds true from maternal age of 0 years till at least 45 years or so. Both the Poisson and logistic regressions curve upwards away from the linear regression at young and old maternal ages due to their decaying exponential nature.

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: The ratio of the odds of delivering a child who is small for gestational age (SGA) based on log2 maternal age was examined using logistic regression. 95% Confidence intervals were calculated using robust standard errors.

Results: Through a logistic regression analysis it was found that every 2-fold increase maternal age is associated with a 48% decrease in the odds of delivery of a SGA infant in the older group compared to the younger group1. 95% confidence intervals suggest that it would not be unusual if the true decrease in the odds of delivery of a SGA infant was between -1% and 60%. A two-sided p value (p=0.05) shows that this observation is not statistically significant at the 0.05 level. Therefore, we cannot reject the null hypothesis that there is no association between maternal age and SGA.

1 The estimate of the slope is -0.66. The exponentiated estimate of the slope is 0.516. Since maternal age is on a base 2 log scale a 1 unit increase in log2(age) is a 2-fold increase in maternal age. So for every 10-fold increase in maternal age we see a 0.516-fold increase in SGA, i.e. a 48% decrease.

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

This is a silly analysis to do since we do not believe that age acts multiplicatively on SGA. It would make more sense if the predictor was another biological variable such as CRP whose distribution follows a log distribution. It is also really hard to interpret. We almost never think of age on a log scale so a 2-fold difference in age is not intuitive.