## Midterm Review Problems

## Epi/Biost 536

## 10.28.13

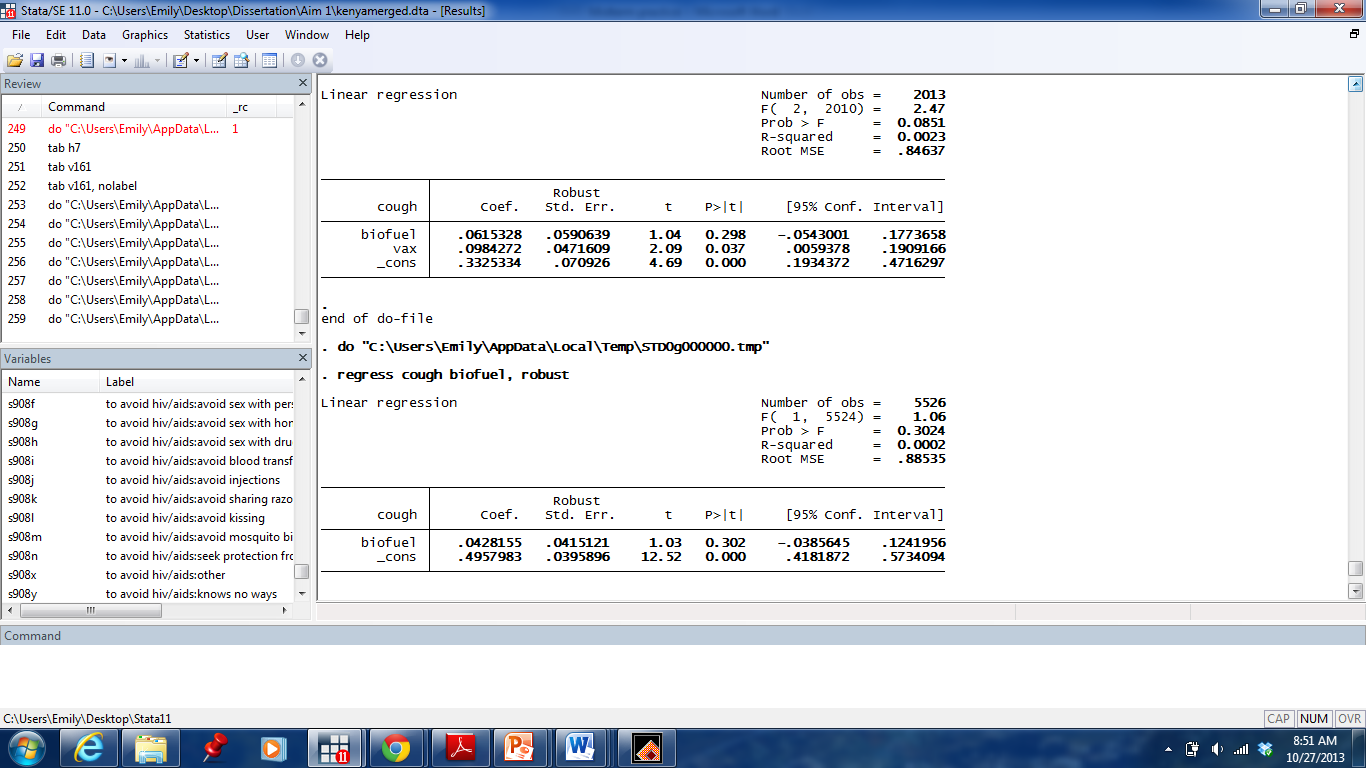
1. A cohort study was conducted in an East African country which included 5526 children under the age of 5 years. The study investigators were interested in the association between childhood pneumonia and use of biofuel. The variables available in the dataset include:
   * mage- maternal age (years)
   * binmage- whether the mother was older than 25 years of age (coded as 1) or younger than or equal to 25 years of age (coded as 0)
   * cough- whether or not the child was ill with a cough in the two weeks before the study visit (0/1)
   * biofuel- whether the mother of the child cooked with charcoal, wood, or animal dung (coded as 1), compared to kerosene or electricity (coded as 0)

Stratified by use of biofuel, the descriptive statistics are as follows:

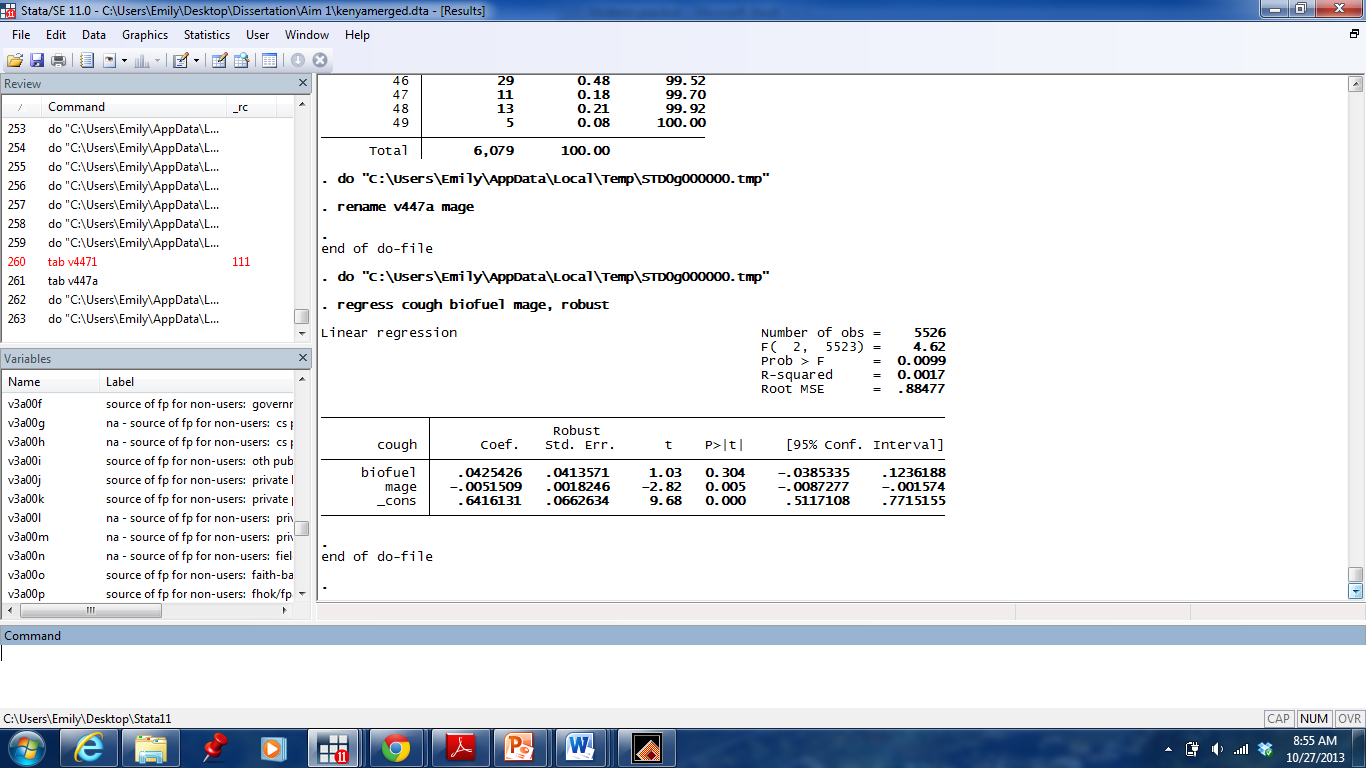
|  |  |  |  |
| --- | --- | --- | --- |
|  | Use Biofuel | Do not use Biofuel | Total |
| Maternal age, mean (sd),years | 28.2 (6.7) | 28.5 (6.0) | 28.3(6.7) |
| Maternal age over 25 years, n(percent) | 3,198 (66.2) | 344 (58.5) | 3,542(59.2) |
| Child with cough, n(percent) | 1,360(26.9) | 118(24.8) | 1,478(26.7) |

The investigators ran a variety of regression models to understand the relationship between “cough” and “biofuel.”

*Model (1)*



*Model (2)*



1. Provide a full interpretation of the coefficient for biofuel in Model 1.

**This is a *risk difference* model.**

**Mention:**

* **Effect size**
* **Statistical significance**

**On average, a 0.043 higher proportion of children whose mothers use biofuel had a cough compared to children whose mothers did not use biofuel. This result is not statistically significant at α=0.05. The observed difference would not be unusual if the true difference in risk is anywhere between -0.039 and .124.**

**OR**

**If you use percentages: Children whose mothers use biofuel have a 4.3% *percentage points* higher risk of a cough, on average, compared to children whose mother’s did not use biofuel.**

1. Provide a full interpretation of the coefficient for biofuel in Model 2.

**On average and among children of mothers of the same age, a 0.043 higher proportion of children whose mothers use biofuel had a cough compared to children whose mothers did not use biofuel. This result is not statistically significant at α=0.05. The observed difference would not be unusual if the true difference in risk is anywhere between -0.039 and .124.**

1. Do these two models provide evidence that maternal age was confounding the relationship between biofuel and cough? Explain briefly.

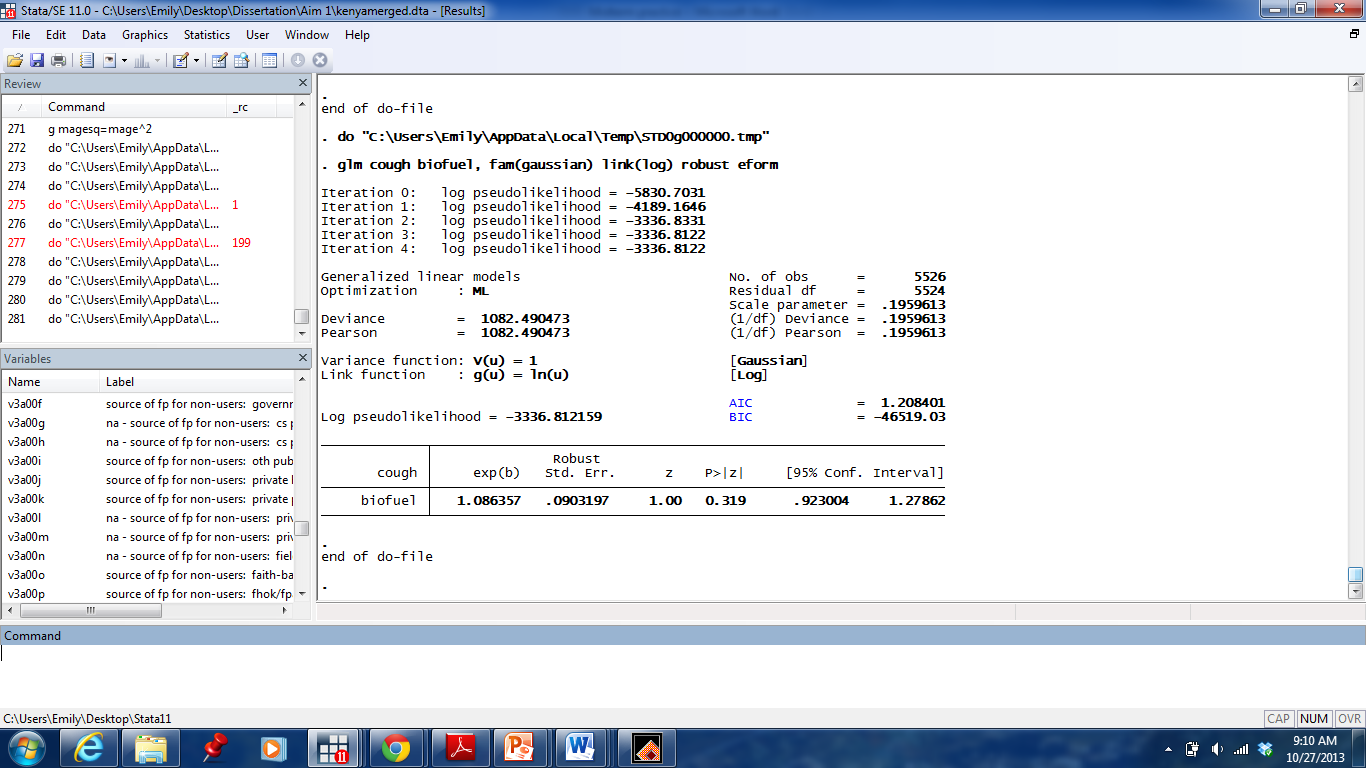
**Mention:**

* **No there is not evidence of confounding.**
* **How we would be able to tell if there were confounding**
* **Why we could use these models to assess symptoms of confounding**

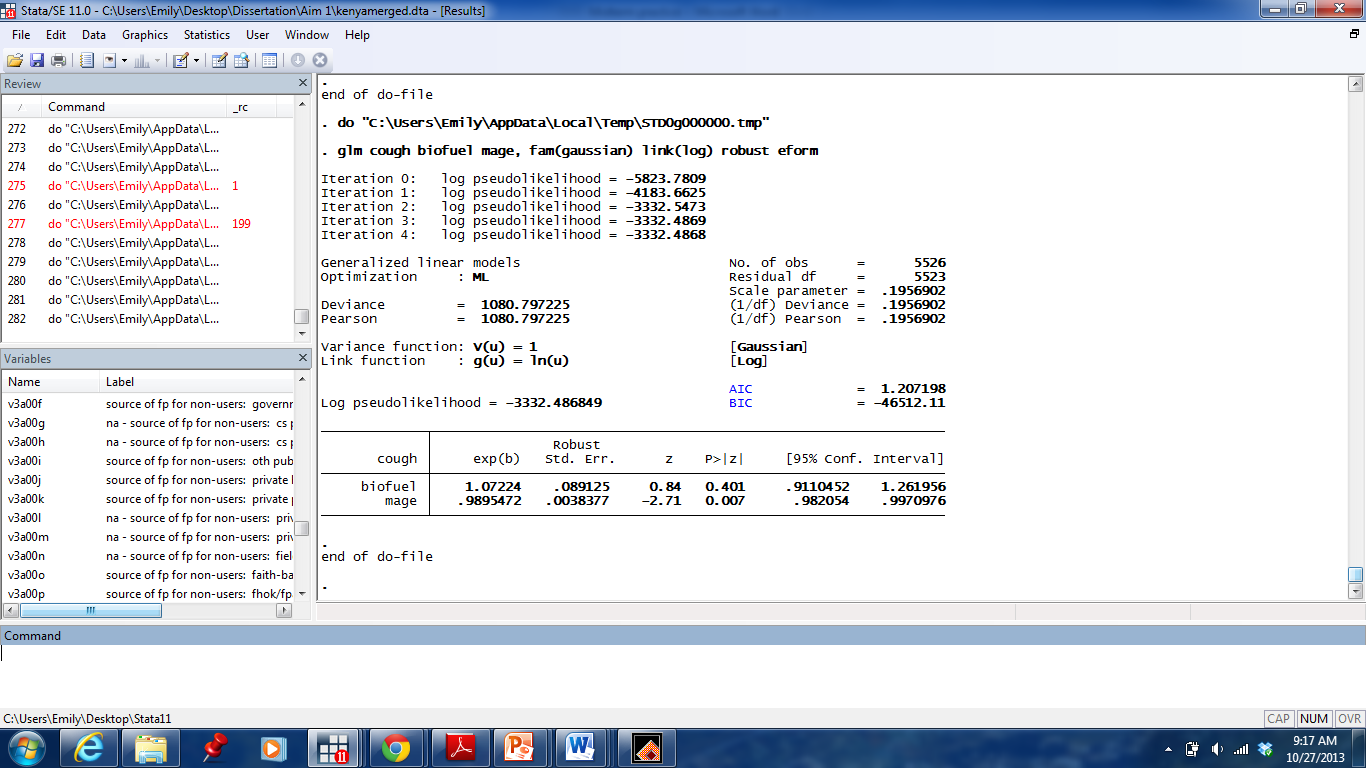
**No. For a linear model, we would expect that confounding would be evidenced by a change in the coefficient for biofuel upon the addition of a term for maternal age. Because the two coefficients in these two models are nearly the same, we do not have evidence for confounding.**

1. The investigators were not satisfied that these two original models were precisely answering their pre-specified scientific question, so they ran FOUR additional regression models:

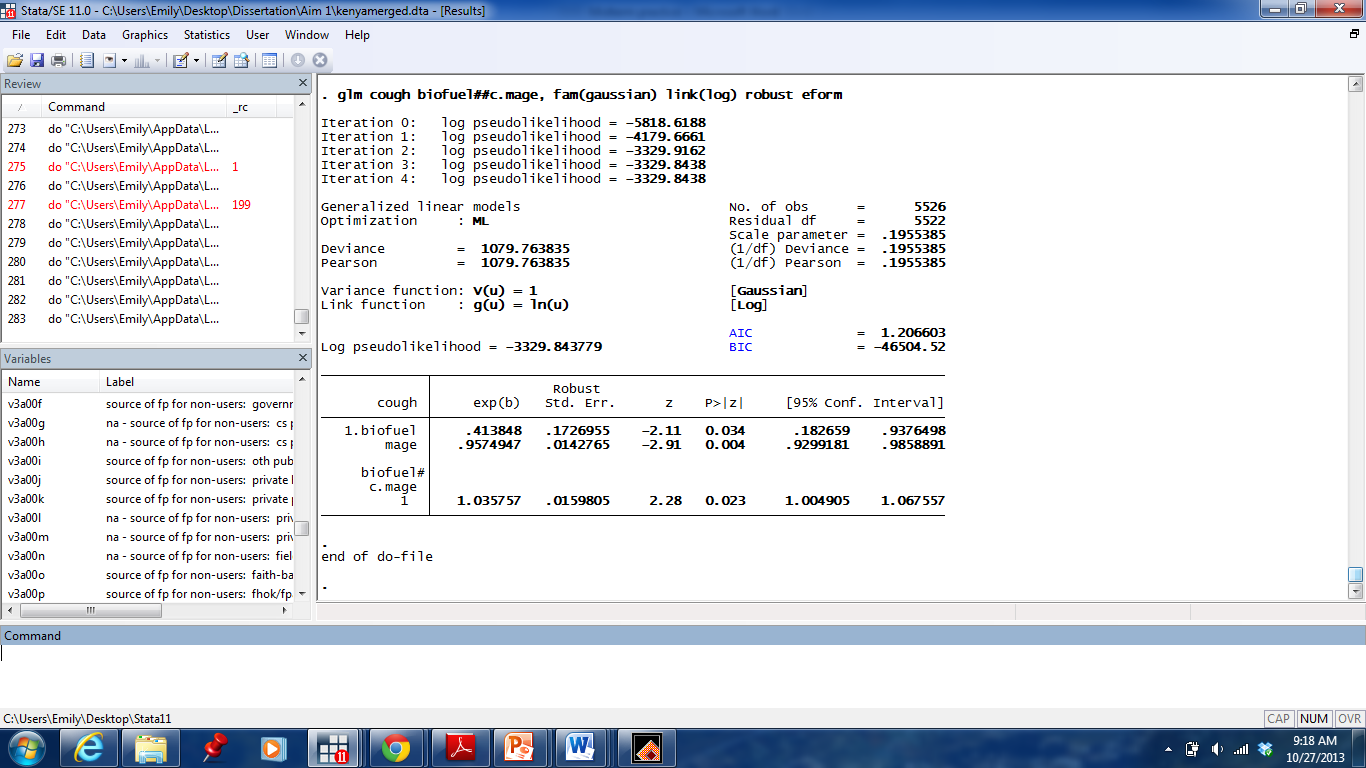
*Model (3)*



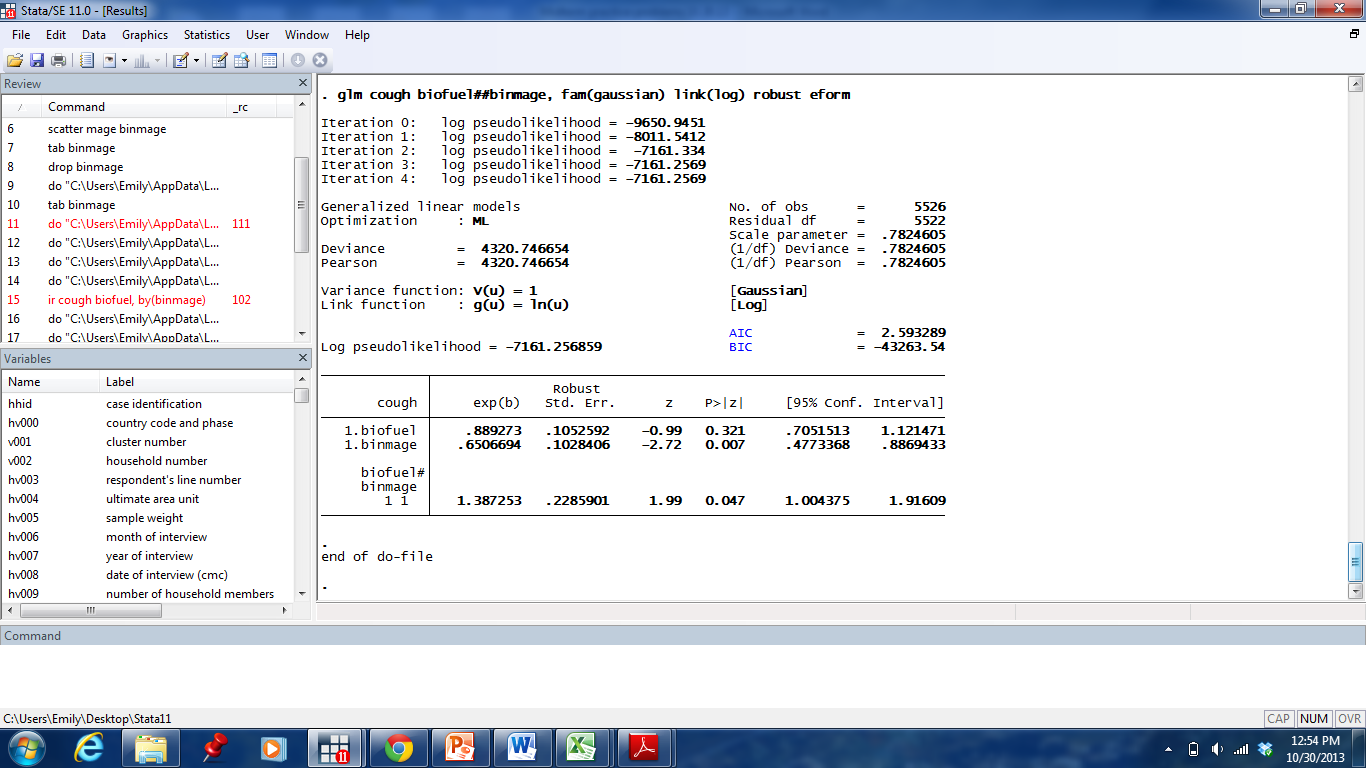
*Model (4)*



*Model (5)*



*Model (6)*



1. Why do you think the investigators did not choose to use a logit link for these four models?

**Mention:**

* **OR does not approximate the RR when the disease is common**
* **RR is easy to understand**
* **RR stays constant regardless of “contamination” of the population with non-susceptible individuls**
* **RR is collapsible across strata**

1. Provide a full interpretation of the coefficient for biofuel in Model 3.

**On average, children whose mother used biofuel were 1.09 times as likely to have a cough compared to those children whose mothers did not use biofuel. This result is not statistically significant at α=0.05. The observed ratio would not be unusual if the true ratio is anywhere between 0.923 and 1.28.**

1. Provide a full interpretation of the coefficient for biofuel in Model 4.

**On average among children with mothers of the same age, children whose mother used biofuel were 1.07 times as likely to have a cough compared to those children whose mothers did not use biofuel. This result is not statistically significant at α=0.05. The observed ratio would not be unusual if the true ratio is anywhere between 0.923 and 1.26.**

1. Do Models 4 and 5 provide evidence that maternal age was confounding the relationship between biofuel and cough? Explain briefly.

**Mention:**

* **Evidence of confounding is not strong but there is a slight difference in the two estimates**
* **This is also a linear model on the log scale, so the estimates are collapsible (we should be able to assess confounding by a change in the estimate)**
* **The slight difference in the estimates might be due to some effect modification on the relative risk scale.**

1. Which of these four models is “saturated”? Why?

**Mention:**

* **Model 6 is saturated**
* **Model 6 estimates a proportion for every group defined**
* **Models 3, 4 and 5 borrow information across groups (mainly maternal age) to create estimates**

1. In Model 5, how would you statistically test the hypothesis that biofuel had a significant influence on cough?

**Mention:**

* **Testing the coefficient of biofuel and the interaction together is required**
* **The “test” command in stata provides a post-estimation Wald test of coefficients or linear combinations (ie, the combination of the coefficient for the intercept and the coefficient for biofuel)**
* **LRT is also possible with GLM**

1. Is there evidence from any of these four models that the effect of biofuel on coughing was different according to maternal age?

**Mention:**

* **Yes there is evidence of effect modification**
* **Both Models 5 and 6 demonstrate this evidence**
* **Model 5 shows that, for each year of age, the effect of biofuel increases**
* **Model 6 shows that, for mothers over the age of 25, the impact of biofuel is higher**

1. If there is effect modification by maternal age, on which scale is it? (RD, RR, OR) Explain briefly.

**Mention:**

* **There is effect modification on the RR scale**
* **It is possible for there to be effect modification on the OR or RD scales, but we did not test for those**

1. In what ways could the interpretation of the inference change if the investigators had used a binomial family for Model 4? Where would you look in the output for the influence of this change in family?

**Mention:**

* **The interpretation of inference probably would not change much**
* **The use of the binomial family only changes the regression weighting used across maternal age and biofuel**
* **The binomial family uses the inverse of the variance to weight: 1/p(1-p), so the highest variance, and lowest weight, would occur where the mean is around 50%**
* **There is the potential for this change in the family to change both the estimate and the standard error slightly**
* **There is potential for the binomial family to be more efficient when it accurately reflects the data**

1. The investigators are concerned about residual confounding by maternal age. Suggest a method by which they could control more thoroughly for age. Explain your choice briefly.

**Mention:**

* **We’ve already used linear continuous and binary**

**Defend pros/ cons of one of the following:**

* **Transformed continuous**
* **Polynomial continuous**
* **Dummy variables**
* **Piecewise linear**
* **Splines**

1. The investigators decided that Model 5 was the most appropriate to answer their pre-specified scientific question. Give an example of a scientific question that Model 5 is answering.

**Eg: Is using biofuel more or less dangerous to the health of children when older mothers use it?**