**Lab 2 - Assignment**

Due 10am Friday, March 7

**Goals:**

* Estimate crude measures of disease frequency and 95% confidence intervals (95% CI) using generalized linear models.
* Estimate crude measures of effect and 95% CI using generalized linear models.
* Assess for potential effect measure modification.

**Assignment breakdown (100 pt):**

* .Rmd file knits properly – 5 pt
* Statement of collaboration – 5 pt
* Tasks and questions (see breakdown below) - 90 pt

Please label all tasks with a heading and answers to questions with a sub-heading and in your Rmarkdown file.

**Instructions**

1. Generate an RMarkdown file and save as Lab2\_[Firstinitial][Lastname].Rmd. For example, Lab2\_CMarkwalter.Rmd.
2. Load biostat3, tidyverse and broom packages **in that order**. Import the kenya\_lab2.csv dataset.
3. First, take a look at the dataset:
   1. Task 1 (10 pt) - Please do the following:
      1. Tabulate the number of children who died and remained alive, stratified by the bord5, male, and education variables.
      2. Find the total number of children and risk of death in each stratum for each variable.
      3. Transcribe the values to Table 1.

Hint: The table has been completed for the magec variable, so you can use this to check your tabulation method.

1. **Use generalized linear models to estimate crude effect measures** on childhood mortality in Kenya:
   1. Task 2 (10 pt) – Please do the following for each of the bord5, male, and education variables separately:
      1. Fit a linear risk model of child mortality.
      2. Use the linear risk model coefficients to find the risk differences (and their confidence intervals) between groups and transcribe to Table 1.
   2. Task 3 (10 pt) - Please do the following for each of the bord5, male, and education variables separately:
      1. Fit a log risk model of child mortality.
      2. Use the log risk model coefficients to calculate the risk ratios between groups and transcribe to Table 1.
   3. Task 4 (10 pt) - Please do the following for each of the bord5, male, and education variables separately:
      1. Fit a logistic regression model of child mortality.
      2. Use the logistic regression model coefficients to calculate the odds ratios between groups and transcribe to Table 1.

Hint: The table has been completed for the magec variable, so you can use this to check your calculation methods.

1. Now we will **explore how variable coding affects our estimates**. We will do this in our dataset with the maternal age variable.
   1. Task 5 (30 pt)– Please do the following:
      1. Fit a linear risk regression model to examine the association between death and maternal age at delivery using magec.
      2. Using the model coefficients, estimate the risks and 95% CI for death associated with a maternal age of 16, 26, and 40 years. Transcribe these values to Table 2.
      3. Estimate the risk differences and 95% CI for death for maternal age of 16 vs 26 (ref) years and 40 vs 26 (ref) years and transcribe these values to Table 2.
      4. Plot the estimated risk of death and 95% confidence intervals predicted by the model. Hints: see class notes from glm day 1. You’ll need multiple geometries (1 for risk and 1 for confidence intervals). Please plot mage on the x axis and risk on the y axis. Note that the risk is derived from your model with magec as the exposure.
      5. Fit a linear risk regression model to examine the association between death and maternal age at delivery as a simple continuous variable (mage)
      6. Using the model coefficients, estimate the risks and 95% CI for death associated with a maternal age of 16, 26, and 40 years. Transcribe these values to Table 2.
      7. Estimate the risk differences and 95% CI for death for maternal age of 16 vs 26 (ref) years and 40 vs 26 (ref) years and transcribe these values to Table 2.
      8. Plot the estimated risk of death and 95% confidence intervals predicted by the model that includes mage (continuous). Hints: see class notes from glm day 1. You’ll need multiple geometries (1 for risk and 1 for confidence intervals). Please plot mage on the x axis and risk on the y axis.
      9. **Question 1: Consider the two plots - describe how the coding of maternal age in the categorical (magec) vs continuous (mage) models influenced the shape of the curve for the association between death and maternal age estimated by each model.**
2. **Effect modification.** We now want to see if the association for birth order (bord5) you estimated in Table 1 differs by the child’s gender (i.e., whether gender acts as an effect modifier of the relationship)
   1. Task 6 (20 pt) – Please do the following:
      1. Tabulate the number of children who died based on bord5 and stratified by male as well as the total number of children in each stratum and transcribe the values to Table 3.
      2. Fit a linear risk model for the 60-month risk of death according to birth order (bord5) and child’s gender (male) with an interaction term.
      3. Using the coefficients from the model, calculate the risk in each stratified category (and the 95% CIs) and transcribe to Table 3.
      4. Using the coefficients from the model, calculate the risk difference and CIs between bord5 categories within each male stratum and transcribe to Table 3.
      5. Fit a linear risk model for the 60-month risk of death according to birth order (bord5) and child’s gender (male) **without** an interaction term.
      6. Perform a likelihood ratio test comparing the model with an interaction term to the model without the interaction term.
      7. **Question 2: Do you observe effect modification by gender on the association between death and birth order? Use your crude estimate, stratum-specific estimates and CIs in Table 3, and the results of the likelihood ratio test to support your answer.**
3. Add a statement of collaboration
4. Knit your RMarkdown document into an .html file. Check to be sure all of the formatting turned out how you expected!
5. Upload your .Rmd, .html, and .docx files as an assignment on Canvas.

**TABLE 1.** Univariable estimated 60-month risks, risk differences (RD), risk ratios (RR), and incidence odds ratios (IOR) and 95% CIs for child mortality in Kenya. DHS data, 2009. Estimated using generalized linear models. All estimates should be to 3 decimal places.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Dead** | **Alive** | **Total** |  |  |  |  |  |  |  |
|  |  | **N** | **N** | **N** | **Risk** | **RD** | **95% CI** | **RR** | **95% CI** | **IOR** | **95% CI** |
| **Birth order**  **(bord5)** | 1-4 (ref) |  |  |  |  | -- | -- | -- | -- | -- | -- |
| 5+ |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **Sex**  **(male)** | Female (ref) |  |  |  |  | -- | -- | -- | -- | -- | -- |
| Male |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **Mother’s education**  **(education)** | Did not attend school (ref) |  |  |  |  | -- | -- | -- | -- | -- | -- |
| Primary school only |  |  |  |  |  |  |  |  |  |  |
| Post-primary education |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **Maternal age**  **(magec)** | <18 years (ref) | 342 | 2007 | 2349 | 0.146 | -- | -- | -- | -- | -- | -- |
| 18-39 years | 1637 | 12689 | 14326 | 0.114 | -0.031 | -0.047, -0.016 | 0.785 | 0.706, 0.876 | 0.757 | 0.668, 0.860 |
| ≥40 years | 38 | 115 | 153 | 0.248 | 0.103 | 0.037, 0.176 | 1.705 | 1.249, 2.246 | 1.939 | 1.306, 2.822 |
|  |  |  |  |  |  |  |  |  |  |  |  |

**Table 2:** Estimated risks, RD and 95% CI for maternal age at delivery from linear risk regression models. All values should be to 3 decimal places

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | 1. magec  (categorical, disjoint indicator vars) | | 2. mage  (simple continuous) | |
|  |  | Est. | 95% CI | Est. | 95% CI |
| Risk | 16 |  |  |  |  |
| 26 |  |  |  |  |
| 40 |  |  |  |  |
| RD | 16 vs. 26 |  |  |  |  |
| 40 vs. 26 |  |  |  |  |

**Table 3**: Model-estimated risks, risk differences (RD) and 95% confidence intervals (CI) for the 60-month risk of child mortality in association with birth order according to child’s sex and the interaction between them**.** All values should be to 3 decimal places.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Deaths  (N) | Total  (N) | Risk | CI | RD | CI |
| **Female** | Birth order 1-4 |  |  |  |  | 0 | -- |
| Birth order 5+ |  |  |  |  |  |  |
| **Male** | Birth order 1-4 |  |  |  |  | 0 | -- |
| Birth order 5+ |  |  |  |  |  |  |