Stat 230 Group HW 6

Group number:

Recorder: Victor Huang

Grader: Ada Wright

Participants: Bridger & Shelby

Due: 10pm Sunday, Oct 31 (boo.)

Grading due: 10pm Monday, Nov 1

Use this Markdown template to answer the questions below. Knit to a Word or pdf doc, OR do this by hand and take a picture, and upload your answers to the group homework forum. Grader comments can be added as comments to this forum post.

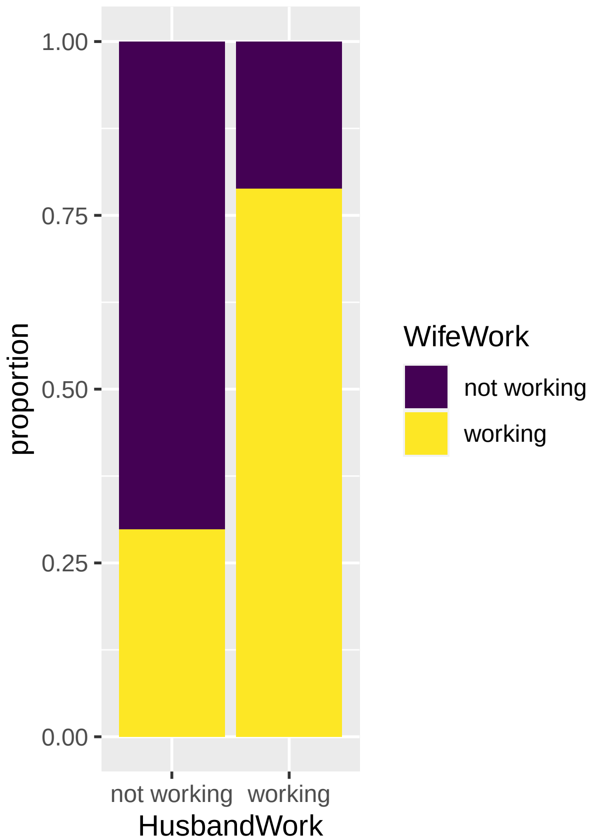
Problem 1: American Community Survey - interpreting logistic model effects

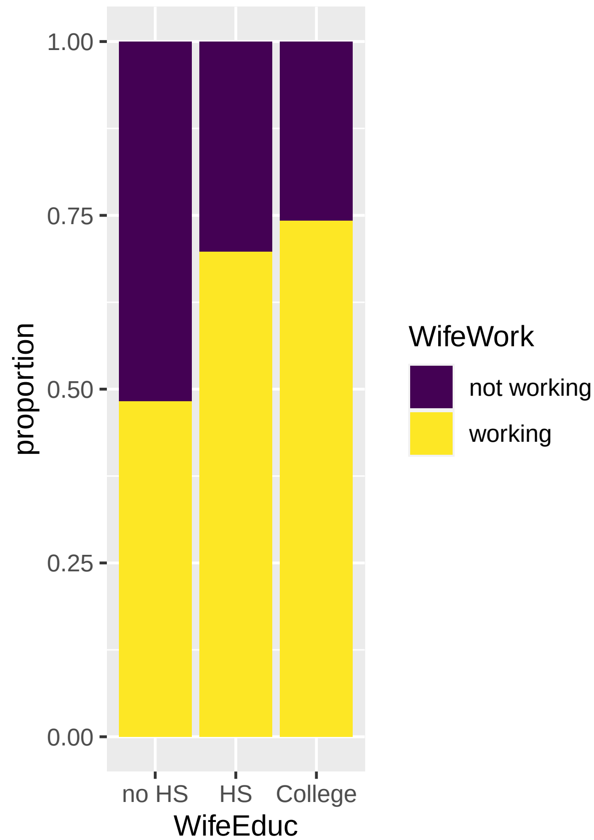
The data for this problem was compiled by Valeria Lambert ('16) and Moira Smith ('16). For married couples in Minnesota, we will model the probability that a wife was employed as a function of their husband's working status, their education level and age. The data comes from a large random sample of MN households obtained in the 2014 American Community Survey ($n = 12,373$). The variables in this data set are:

- `WifeWorkingStatus`: `NotWorking` and `Working` (response)
- `WifeEduc`: `noHS`, `HS`, and `College`
- `WifeAge`: age in years (quantitative)
- `HusbandWorking`: `NotWorking` and `Working`

No R work is needed for this problem. Feel free to write up answers on paper and upload an image to Moodle for your HW submission.

(a) The two stacked bar graphs below show the relationship between wife working status and the two categorical predictors husband's working status and wife's education level. For each bargraph, describe the relationship shown in the plot.





answer: For the first bargraph we see that approximately when the husband is not working, a proportion of 0.3 of the wives work while a proportion of 0.7 of wives do not work. When the husband is working, approximately a proportion of 0.8 of the wives are also working while only a proportion of 0.2 of the wives are not working. For the bar graph on the right, the wives who have no high school education, approximately half of wives are working and the other are not. The working wives proportion increases as education level increases with high school level education wives having a 0.7 working proportion and college level education wives with a 0.75 working proportion.

(b) The R output below shows the summary output for the fitted logistic regression of wife's working status on husband's working status, wife's education level, and wife's age (linear and quadratic effects). Note that the second level of `WifeWork` is `Working`, so `Working` is a "success" in this model. Use these results to estimate the *odds* that a 30-year old wife is working if she has a high school (HS) education and a husband who is also working.

```
> my.glm <- glm(WifeWorkingStatus ~ HusbandWorking+WifeEduc + WifeAge + I(WifeAge^2),
                data = data, family = binomial)
> summary(my.glm)

Call:
glm(formula = WifeWorkingStatus ~ HusbandWorking + WifeEduc +
    WifeAge + I(WifeAge^2), family = binomial, data = data)

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)           -4.8967688  0.3069392 -15.954  < 2e-16 ***
HusbandWorkingWorking  0.7893083  0.0587821  13.428  < 2e-16 ***
WifeEducHS             0.3696654  0.0571197   6.472 9.69e-11 ***
```

```
WifeEducCollege        0.4841689  0.0609396   7.945 1.94e-15 ***
WifeAge                0.2803366  0.0128555  21.807  < 2e-16 ***
I(WifeAge^2)          -0.0033774  0.0001308 -25.830  < 2e-16 ***
```

answer: $\hat{\eta}(working) = -4.8967688 + 0.7893083(1) + 0.3696654(1) + 0.4841689(0) + 0.2803366(30) - 0.0033774(30)^2 = 1.632643$

$odds = e^{1.632643} = 5.117$

The odds for a wife with these conditions to be working is 5.117 (for every single non working wife in these conditions, there will be 5.117 wives working)

(c) Estimate the probability that a 30-year old wife is working if she has a high school (HS) education and a husband who is also working. *answer:* $probability = \frac{e^{1.632643}}{1+e^{1.632643}} = 0.8365$. We get a probability of 0.8365

(d) Using the model in part (b), we see that $\hat{\beta}_{husbandworking} = 0.7893083$. Use this estimated effect to explain the effect of husband's working status on the odds that he has a working wife, holding education and age constant. Does this effect agree with your EDA in part (a)? *answer:* As a binary variable, if the husband is working, it effects the odds of the working wife by $e^{0.7893083}$. If the husband is not working, this term is 0 and there is no effect. This agrees with our EDA initially as we can see a working husband also has a higher odds for having a working wife as well.

(e) Using the model in part (b), we see that $\hat{\beta}_{WifeEducHS} = 0.3696654$. Carefully(!) interpret this effect on the odds of being a working wife. Repeat for $\hat{\beta}_{WifeEducCollege} = 0.4841689$. Do these effects agree with your EDA in part (a)? *answer:* As a binary variable, if the wife is HS educated, it effects the odds of the working wife by $e^{0.3696654}$. If the wife is college education, it effects the odds of the working wife by $e^{0.4841689}$ This agrees with our EDA initially as we can see a higher educated wife also has higher odds for working as well.

(f) Using the model in part (b), how do the odds of working change as a woman increases in age from 30 to 31 years, everything else held constant? Repeat this question for a woman who increases in age from 60 to 61 years. *answer:* While holding everything else constant, increasing a woman's age from 30 to 30 increases the likelihood that she is working by 7.7%. increasing a woman's age from 60 to 61 decreases the likelihood that she is working by 12.0%.

$$probability = \frac{prob@age30}{prob@age30+1} = e^{0.2803366-0.0033774*(2*30+1)} = e^{0.0743} = 1.077$$

$$probability = \frac{prob@age60}{prob@age60+1} = e^{0.2803366-0.0033774*(2*60+1)} = e^{-0.1283} = 0.880$$