American Youth Smoking Rate

Yeyunfei Jiang (1004170354)

2020/10/25

Introduction:

Cigarette smoking amongst children is known to be more common for males than females,in rural areas than urban areas, and to vary by age and ethnicity. The research question of interest ralate to the variation amongst the US states exist, and that there is variation from one school to the next, A secondary task is to convey the differences in the effect of age on smoking for white, Black,and Hispanic americans. The effect is expected to be different by sex and by rurality. Two hypothesis are made as follows:

- Geographic variation (between states) in the rate of students smoking cigarettes issubstantially greater than variation amongst schools. As a result, tobacco controlprograms should target the states with the most smoking and not concern themselveswith finding particular schools where smoking is a problem.
- Rural-urban differences are much greater than differences between states.

The dataset used in this study is from 2014 American National Youth Tobacco Survey, which contains over 18000 survey from Americian youth of different races and schools between age 11 - 18.

Method

The bayesian inference is as follows:

$$Y_{ijk} \sim Bern(\mu_{ijk})$$

$$log(\frac{\mu_{ijk}}{1 - \mu_{ijk}}) = X_{ijk}\beta + U_i + Z_{ij}$$

$$U_i \sim N(0, \sigma_U^2)$$

$$Z_{ij} \sim N(0, \sigma_Z^2)$$

Priors:

$$\beta \sim N(0, 100^2)$$

$$\sigma_U \sim pc.prec(log(2.5)/1.3, 0.5)$$

$$\sigma_Z \sim pc.prec(log(1.15)/1.3, 0.5)$$

Where μ_{ij} is the probability of the person k in j school in i state smoke cigarettes. Y_{ijk} is zero if the kth person hasn't smoke cigarettes. Response variables include age(11-17), sex(levels male, female), race(levels white, black, hispanic, asian, pacific, native), and rural/urban(levels rural, urban). U_i is the state specific random effect, Z_{ij} is the school specific random effect.

The priors for σ_U, σ_Z are exponencial distributed, with median around 0.7,0.11 correspondingly. By the prior informations that collaborating scientists provided:

- Some states having double or triple the rate of smoking update compared other states for comparable individuals.Within a given state.

- Within a given state, a differences of 10% to 20% in smoking rates is moretypical.

By setting the prior of σ_U to be pc.prec(log(2.5)/1.3,0.5). It assumes that person in the 75th percentile of the smoking state has 2.5 times the odds of being smoking cigarettes as the person in the 25th percentile state.

By setting the prior of σ_Z to be pc.prec(log(1.15)/1.3,0.5). It assumes that person in the 75th percentile of the smoking school has 1.15 times the odds of being smoking cigarettes as the person in the 25th percentile school.

Result:

Table 1: Log odds of fixed effect parameter estimate and 95% CI(interaction not included)

|	mean	sd	0.025quant	0.5quant	0.975quant	mode
(Intercept)	-4.078	0.515	-5.186	-4.042	-3.165	-3.968
SexF	-0.182	0.038	-0.257	-0.182	-0.107	-0.182
RuralUrbanRural	1.533	0.581	0.476	1.503	2.756	1.443
ageFac12	0.864	0.553	-0.132	0.832	2.040	0.766
ageFac13	1.634	0.524	0.702	1.599	2.761	1.527
ageFac14	2.047	0.523	1.117	2.012	3.172	1.939
ageFac15	2.892	0.523	1.962	2.857	4.016	2.784
ageFac16	3.220	0.522	2.291	3.184	4.343	3.112
ageFac17	3.767	0.521	2.840	3.732	4.888	3.659
ageFac18	3.683	0.528	2.743	3.648	4.815	3.576
Raceblack	0.281	0.879	-1.567	0.324	1.890	0.410
Racehispanic	0.331	0.683	-0.983	0.322	1.698	0.304
Raceasian	0.078	1.133	-2.406	0.176	2.033	0.383
Racenative	1.326	1.151	-1.191	1.422	3.321	1.625
Racepacific	-43.031	20.796	-92.220	-39.180	-13.760	-27.169

Table 2: Probability of fixed effect parameter estimate and 95% CI(interaction not included)

	mean	sd	0.025quant	0.5quant	0.975quant	mode
(Intercept)	0.017	0.626	0.006	0.017	0.041	0.019
SexF	0.455	0.510	0.436	0.455	0.473	0.455
RuralUrbanRural	0.822	0.641	0.617	0.818	0.940	0.809
ageFac12	0.704	0.635	0.467	0.697	0.885	0.683
ageFac13	0.837	0.628	0.669	0.832	0.941	0.822
ageFac14	0.886	0.628	0.753	0.882	0.960	0.874
ageFac15	0.947	0.628	0.877	0.946	0.982	0.942
ageFac16	0.962	0.628	0.908	0.960	0.987	0.957
ageFac17	0.977	0.627	0.945	0.977	0.993	0.975
ageFac18	0.975	0.629	0.940	0.975	0.992	0.973
Raceblack	0.570	0.707	0.173	0.580	0.869	0.601
Racehispanic	0.582	0.665	0.272	0.580	0.845	0.575
Raceasian	0.519	0.756	0.083	0.544	0.884	0.595
Racenative	0.790	0.760	0.233	0.806	0.965	0.835
Racepacific	0.000	1.000	0.000	0.000	0.000	0.000

Table 3: Random effect sandard diviation and 95% CI

	mean	sd	0.025quant	0.5quant	0.975quant	mode
SD for state	0.271	0.080	0.135	0.265	0.448	0.249
SD for school	0.478	0.042	0.399	0.476	0.566	0.473

Prior and posterior distribution graph for school, classUnique, and studentUnique:
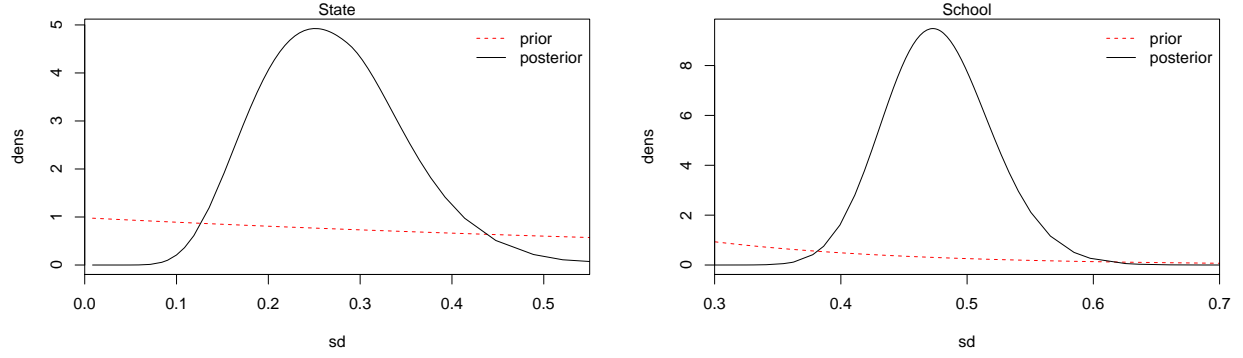
Posterior medians of probabilities of smoking over age by race:
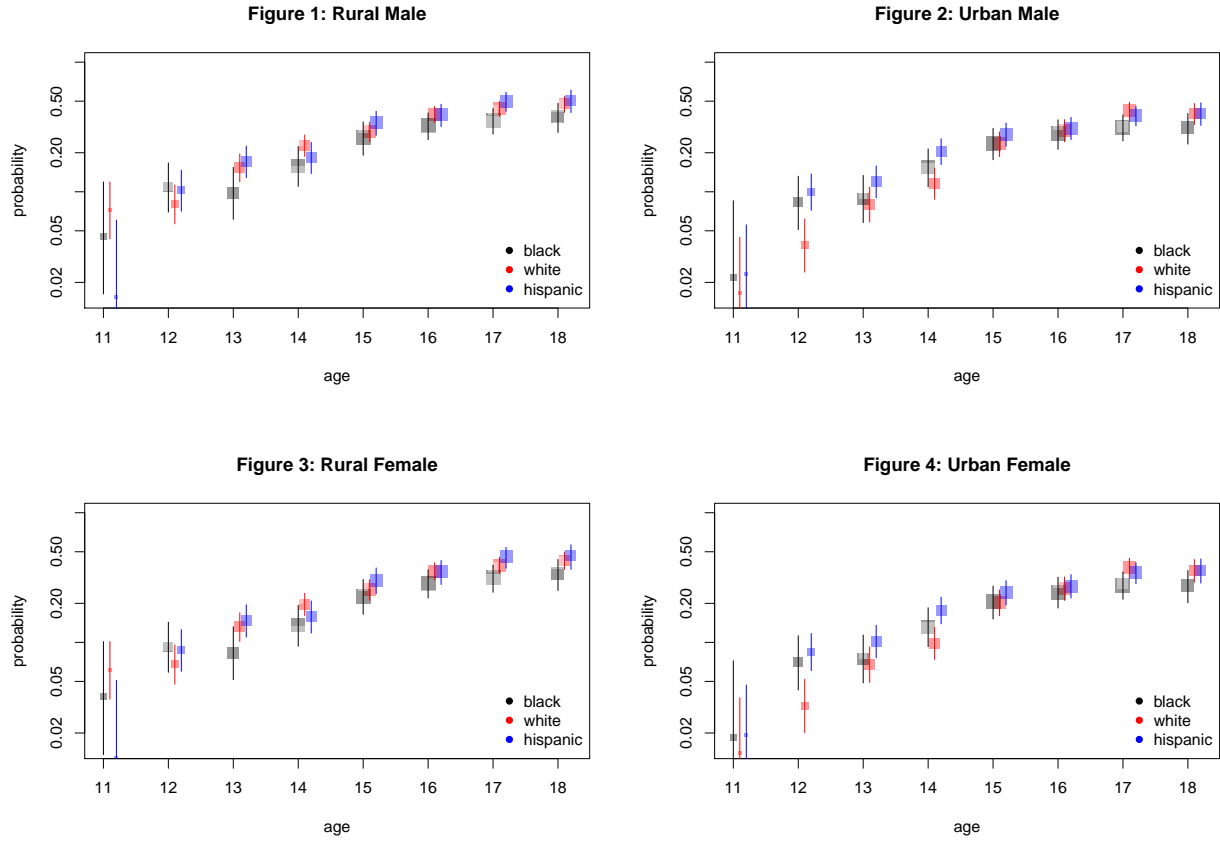
Discussion:

As shown in Table 5, a 95% credible interval for standard deviation of state specific random effect is (0.135, 0.448) , and school specific random effect is (0.399, 0.566). Apparently, the credible interval overlaps with each other, which does not support the hypothesis that geographic variation in the rate of students smoking cigarettes is substantially greater than variation amongst schools. And also, by looking at the median (0.265 and 0.476), variation amongst school should be greater than variation amongst states.

From Table 5 , the 95th CI of odds of the person in the smoking state with one standard deviation compare to in a typical state(mean) is exp(0.135, 0.448) = (1.145, 1.565). And from Table 3, the 95th CI of odds of the person in Rural compare to in urban is exp(0.476, 2.756) = (1.609, 15.737). By comparing these two CIs, we can see that they do not overlap, and the effect of RuralUrban is larger than the effect of state. Thus data support that Rural-urban differences are greater than differences between states.

Figures (1-4) showing the posterior medians of probabilities of smoking over age by black, white, and hispanic americans. From the trending, comparing figure 1 to figure 3, and figure 2 to figure 4. For the same urban/rural status, the preobability change by the sec in each age group for all three races are about constant. By comparing figure 1 to figure 2, and figure 3 to figure 4, we can see that the probability of white and hispanic americans of smoking cigarettes is low when age is around 11 -14 in urban area, and after 14, they starts to catch up. While in rural area, they have a higher probability of smoking at young age, and have a more smooth increase over age compared to urban area. The black americans, on the other side, seems to have a similar growth of probability of smoking as age growth under the effect of urban/rural.

Summary of Conclusion:

The study modelled the relation between each age, gender, race, and urban/rural with probability of smoking cigarettes among teenagers within different schools and states. It counts the specific random effect of school, and state, and as a result, the data does not support the hypothesis of geographic variation in the rate of students smoking cigarettes is substantially greater than variation amongst schools, but also put forward that variation amongst school has a higher probability of been greater than variation amongst states.

When considering the effect of urban/rural, this research find out that a person in rural area is around (1.609, 15.737) times the odds of being smoking cigarettes as the person in urban area. While a person in the smoking state with one standard deviation is around (1.145, 1.565) times the odds of being smoking cigarettes as the person in a typical state. Hence the study suggest that Rural-urban differences are greater than differences between states.

It is also crucial that the reserach finds out that age effect on smoking for white and hispanic americans are different by rurality, but not sex. First of all, the effect of age on smoking acts on all three races would been decreased as they are older (age 17-18). The effect of age on white and hispanic americans of smoking cigarettes is low when age is around 11 -14 in urban area, in contrary, the effect is more efficient when they are on ages of 14-18 comparing to black americans. While in rural area, they have a higher probability of smoking at young age, and have a more smooth increase over age compared to urban area. The black americans, on the other side, seems to have a similar growth of probability of smoking as age growth under the effect of urban/rural. For different sex, nor three of them would impact on the age effect on smoking.