SDHRM
Homework 3

Part A: Ecologic Data (*5 points*)

Review the SAS output for the best zip-level, tract-level, and individual-level linear regression models and answer the questions following each model. Immediately below are the hypothesis, the study design, the list of variables included in the models, and the questions used to create the social support score.

Hypothesis: Among participants in the Cardiovascular Health Study, increased social support as measured by the social support score is associated with decreased body mass index (BMI)

Study design: Cross-sectional study of elderly adults in the Cardiovascular Health Study

Variable	Label	N	Min	Median	Max
scorebl_zip	mean social support score at zip level	414	6.00	8.00	17.00
male_zip	proportion male at zip level	414	0.00	0.38	1.00
agebl_zip	mean age at zip level	414	65.00	72.72	95.00
white_zip	proportion white at zip level	414	0.00	1.00	1.00
smoker_zip	proportion smokers at zip level	414	0.00	0.00	1.00
bpbl_zip	proportion with BP history at zip level	414	0.00	0.50	1.00
poorhealth_zip	proportion with poor health at zip level	414	0.00	0.00	1.00
married_zip	proportion married at zip level	414	0.00	0.80	1.00
bmi_zip	mean BMI at zip level	414	15.79	26.49	46.53
scorebl_tract	mean social support score at tract level	1024	6.00	8.00	17.00
male_tract	proportion male at tract level	1024	0.00	0.33	1.00
agebl_tract	mean age at tract level	1024	64.00	72.50	95.00
white_tract	proportion white at tract level	1024	0.00	1.00	1.00
smoker_tract	proportion smokers at tract level	1024	0.00	0.00	1.00
bpbl_tract	proportion with BP history at tract level	1024	0.00	0.50	1.00
poorhealth_tract	proportion with poor health at tract level	1024	0.00	0.00	1.00
married_tract	proportion married at tract level	1024	0.00	0.72	1.00
bmi_tract	mean BMI at tract level	1024	14.85	26.41	46.53
scorebl	social support score	5274	6.00	7.00	22.00
male	male sex	5274	0.00	0.00	1.00
agebl	calculated age at baseline	5274	63.00	71.00	100.00
white	white race	5274	0.00	1.00	1.00
smoker	current smoker	5274	0.00	0.00	1.00
bpbl	doctor told you had high blood pressure	5274	0.00	0.00	1.00
poorhealth	poor health	5274	0.00	0.00	1.00
married	married	5274	0.00	1.00	1.00
bmi	body mass index	5274	14.68	26.15	58.91

Social support score (calculated as sum of questions 1-8):

1. How many relatives do you see or hear from at least once per month?
2. How often do you see or hear from the relative with whom you have the most contact?
3. How many relatives do you feel close to?
4. How many close friends do you have?
5. How many of these friends do you see or hear from at least once per month?
6. How often do you see or hear from the friend with whom you have the most contact?
7. When you have an important decision to make, how often do you have someone you can talk to about it?
8. When other people you know have an important decision to make, how often do they talk to you about it?

Best zip-level model:

The REG Procedure

Dependent Variable: bmi_zip mean BMI at zip level

Number of Observations Read 414

Number of Observations Used 414

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	776.09824	86.23314	6.06	<.0001
Error	404	5749.86321	14.23233		
Corrected Total	413	6525.96145			

Root MSE	3.77258	R-Square	0.1189
Dependent Mean	26.64374	Adj R-Sq	0.0993
Coeff Var	14.15933		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	37.08595	2.95113	12.57	<.0001
scorebl_zip	mean social support score at zip level	1	-0.19229	0.08923	-2.16	0.0317
male_zip	proportion male at zip level	1	0.63077	0.53042	1.19	0.2351
agebl_zip	mean age at zip level	1	-0.12120	0.03651	-3.32	0.0010
white_zip	proportion white at zip level	1	-0.78600	0.61409	-1.28	0.2013
smoker_zip	proportion smokers at zip level	1	-1.97767	0.71837	-2.75	0.0062
bpbl_zip	proportion with BP history at zip level	1	2.14656	0.45164	4.75	<.0001
poorhealth_zip	proportion with poor health at zip level	1	0.48757	0.52996	0.92	0.3581

Question A.1: Succinctly interpret the point estimate for the “scorebl_zip” parameter for this linear regression model. *(1 point)*

Best tract-level model:

The REG Procedure
Model: MODEL1
Dependent Variable: bmi_tract mean BMI at tract level

Number of Observations Read	1024
Number of Observations Used	1024

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	1801.80769	200.20085	13.55	<.0001
Error	1014	14982	14.77468		
Corrected Total	1023	16783			

Root MSE	3.84378	R-Square	0.1074
Dependent Mean	26.69485	Adj R-Sq	0.0994
Coeff Var	14.39897		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	37.84222	1.85138	20.44	<.0001
scorebl_tract	mean social support score at tract level	1	-0.13020	0.05793	-2.25	0.0248
male_tract	proportion male at tract level	1	-0.15521	0.34663	-0.45	0.6544
agebl_tract	mean age at tract level	1	-0.12784	0.02360	-5.42	<.0001
white_tract	proportion white at tract level	1	-1.38011	0.33929	-4.07	<.0001
smoker_tract	proportion smokers at tract level	1	-2.07781	0.43475	-4.78	<.0001
bpbl_tract	proportion with BP history at tract level	1	1.67877	0.29805	5.63	<.0001
poorhealth_tract	proportion with poor health at	1	0.28561	0.33512	0.85	0.3943

Question A.2: Succinctly interpret the point estimate for the “scorebl_tract” parameter for this linear regression model. Why does it differ from the “scorebl_zip” parameter in the model above? *(1 point)*

Best individual level model:

The REG Procedure
Model: MODEL1
Dependent Variable: bmi body mass index

Number of Observations Read 5274
Number of Observations Used 5274

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	12307	1538.34588	76.91	<.0001
Error	5265	105316	20.00297		
Corrected Total	5273	117622			

Root MSE	4.47247	R-Square	0.1046
Dependent Mean	26.74663	Adj R-Sq	0.1033
Coeff Var	16.72161		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	38.15101	0.91023	41.91	<.0001
scorebl	social support score	1	-0.06808	0.02549	-2.67	0.0076
male	male sex	1	-0.15823	0.13258	-1.19	0.2327
agebl	calculated age at baseline	1	-0.13453	0.01173	-11.47	<.0001
white	white race	1	-1.93250	0.17208	-11.23	<.0001
smoker	current smoker	1	-2.10129	0.19339	-10.87	<.0001
bpbl	doctor told you had high blood pressure	1	1.53716	0.12642	12.16	<.0001
poorhealth	poor health	1	0.63900	0.14708	4.34	<.0001
married	married	1	-0.09487	0.14553	-0.65	0.5145

Question A.3: Succinctly interpret the point estimate for the “scorebl” parameter for this linear regression model. Why does it differ from the “scorebl_tract” parameter in the model above? *(1 point)*

Question A.4: Which model is more likely to reflect the true relationship between social support and BMI? Why? *(2 points)*

Part B: Mulit-level Data (*5 points*)
Review the SAS output for the ordinary least squared and mixed model output below and answer the 3 questions at the end. Immediately below are the hypothesis, the study design, and the list of variables included in the models.

Hypothesis: Census tracts where the average amount of alcohol consumption is higher are more likely to have residents with increased systolic blood pressure compared to census tracts where the average alcohol consumption is lower in residents after controlling for individual-level alcohol consumption.

Study design: Cross-sectional study of 5701 elderly adults living in 1098 unique census tracts who were participants in the Cardiovascular Health Study

Variable	Label	N	Min	Median	Max
male	male sex	5701	0.00	0.00	1.00
agebl	calculated age at baseline	5701	63.00	72.00	100.00
white	white race	5701	0.00	1.00	1.00
bpbl	doctor told you had high blood pressure	5701	0.00	0.00	1.00
bmi	body mass index	5701	14.68	26.15	58.91
alcohbl	total alcohol per wk	5701	0.00	0.00	85.00
sysbl	systolic blood pressure	5701	77.00	134.00	234.77
white_tract	proportion white at tract level	5701	0.00	0.99	1.00
alcohbl_tract	mean alcohol at tract level	5701	0.00	1.92	42.00
alcohbl_balanced	individual mean-balanced alcohol	5701	-41.75	-0.02	84.98

Ordinary linear regression model:

The REG Procedure
Dependent Variable: sysbl systolic blood pressure

Number of Observations Read 5701
Number of Observations Used 5701

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	382160	47770	116.97	<.0001
Error	5692	2324668	408.40974		
Corrected Total	5700	2706828			

Root MSE	20.20915	R-Square	0.1412
Dependent Mean	136.35119	Adj R-Sq	0.1400
Coeff Var	14.82140		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	69.60539	4.41393	15.77	<.0001
agebl	calculated age at baseline	1	0.78595	0.04859	16.17	<.0001
male	male sex	1	-1.46150	0.55180	-2.65	0.0081
white	white race	1	-3.48176	0.74070	-4.70	<.0001
bmi	body mass index	1	0.20973	0.05930	3.54	0.0004
alcohbl_balanced	individual alcohol consumption minus tract level mean	1	0.19150	0.04293	4.46	<.0001
bpbl	doctor told you had high blood pressure	1	12.66129	0.55215	22.93	<.0001
white_tract	proportion white at tract level	1	0.99768	1.47777	0.68	0.4996
alcohbl_tract	mean alcohol at tract level	1	0.24746	0.11317	2.19	0.0288

Multilevel mixed model:

The Mixed Procedure
Model Information

Data Set	WORK.PERSON_TRACT
Dependent Variable	sysbl
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
tract_unique	1098	147500003 202101408 263200048 [..] 9960300017

Number of Observations

Number of Observations Read	5701
Number of Observations Used	5701
Number of Observations Not Used	0

Covariance Parameter Estimates

Cov Parm	Estimate
tract_unique	10.3357
Residual	397.61

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	67.8702	4.4335	1097	15.31	<.0001
agebl	0.7986	0.04881	4595	16.36	<.0001
male	-1.4312	0.5484	4595	-2.61	0.0091
white	-2.7641	0.7927	4595	-3.49	0.0005
bmi	0.2174	0.05919	4595	3.67	0.0002
alcohbl_balanced	0.1778	0.04298	4595	4.14	<.0001
bpbl	12.6889	0.5498	4595	23.08	<.0001
white_tract	1.4692	1.4820	4595	0.99	0.3216
alcohbl_tract	0.2030	0.1139	4595	1.78	0.0748

B.1. Why are the parameter estimates and the standard errors for the ALCOHBL_TRACT different in each model? *(1 point)*

B.2. Which model is preferable and why? *(2 points)*

B.3. Can you quantify the amount of tract-to-tract variation in systolic blood compared to person-to-person variation? Where does most of the variation lie? *(2 points)*

Version 2/15/10