Travel Demand Analysis
Exercise 3

310702051 葉家榮

2022-04-20

This assignment is also uploaded in the following website, for detailed information and code. https://chiajung-yeh.github.io/Travel-Demand-Analysis/multilevel-model.html

Problem

You are provided with a travel survey data of Tainan (from "南臺區域運輸規劃"). Use trip distance as the outcome variable, and fit multilevel models with the grouping structure, respondents within districts.

1. Fit two constant-only models, one using the original value of trip distance and the other transforming the value via logarithm. Explain the results, compare the two models, and pick one for the next step.

2. Develop fixed-effect only (varying intercept) models; at least two respondent level and one district-level variables need to be included in the models. You need to find your own district-level variables. Explain your models, compared them, and pick one for the next step.

Constant-only Models

The result of two constant-only models are shown in Table 1. Note that the table provided by package `sjPlot`, which gives a more detailed information of the model result, is shown **here**. The original value of trip distance is selected to be dependent variable in the first model (Model 1), while the second model (Model 2) takes the logarithm of trip distance. The trip distance is transformed to "kilometers" in advance, to avoid a huge coefficient in the result. The intercept of Model 1 is 5.328, which is the fixed effect of the constant term. The random effect for each district represents how much the intercept is shifted up or down in particular district. The random effect of each district is shown in the left panel of Figure 1. The red line implies that the estimates of trip distance is lower than the fixed effect, while the blue one is higher than that. Take 龍崎區 for instance, the random effect of it is 5.513 (check out by using `ranef()`, or the toppest point in Figure 1). Since the total effect is composed of the fixed effect and the random effect, the final intercept would be $5.328 + 5.513 = 10.841$. Note that if the line intersects with the "x=0" vertical line, it means that the estimate of trip distance for that district is not significantly inconsistent with the fixed effect, namely, the random effect might not be true.

The fixed intercept of Model 2 is 0.835, and the random effect is illustrated in the right panel of Figure 1. We can find that when taking the logarithm form of trip distance, the sign of some random effects are not remained the same compared to the previous model. To choose a model, it is not correct to simply observe the value of AIC, BIC or log likelihood, since the dependent variable of two models are not the same. In terms of Intraclass Correlation Coefficient (ICC), it is the correlation between two observations within the same group. It is calculated as: $\frac{\tau_0}{\tau_0+\sigma^2}$, where τ_0 is the between-group variation, while σ^2 is the within-group

Table 1: Constant-only Models

	Model 1	Model 2
(Intercept)	5.328***	0.835***
	(0.421)	(0.055)
AIC	21859.757	11402.361
BIC	21878.163	11420.767
Log Likelihood	−10926.878	−5698.181
Num. obs.	3413	3413
Num. groups: district	37	37
Var: district (Intercept)	5.653	0.079
Var: Residual	34.446	1.624

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

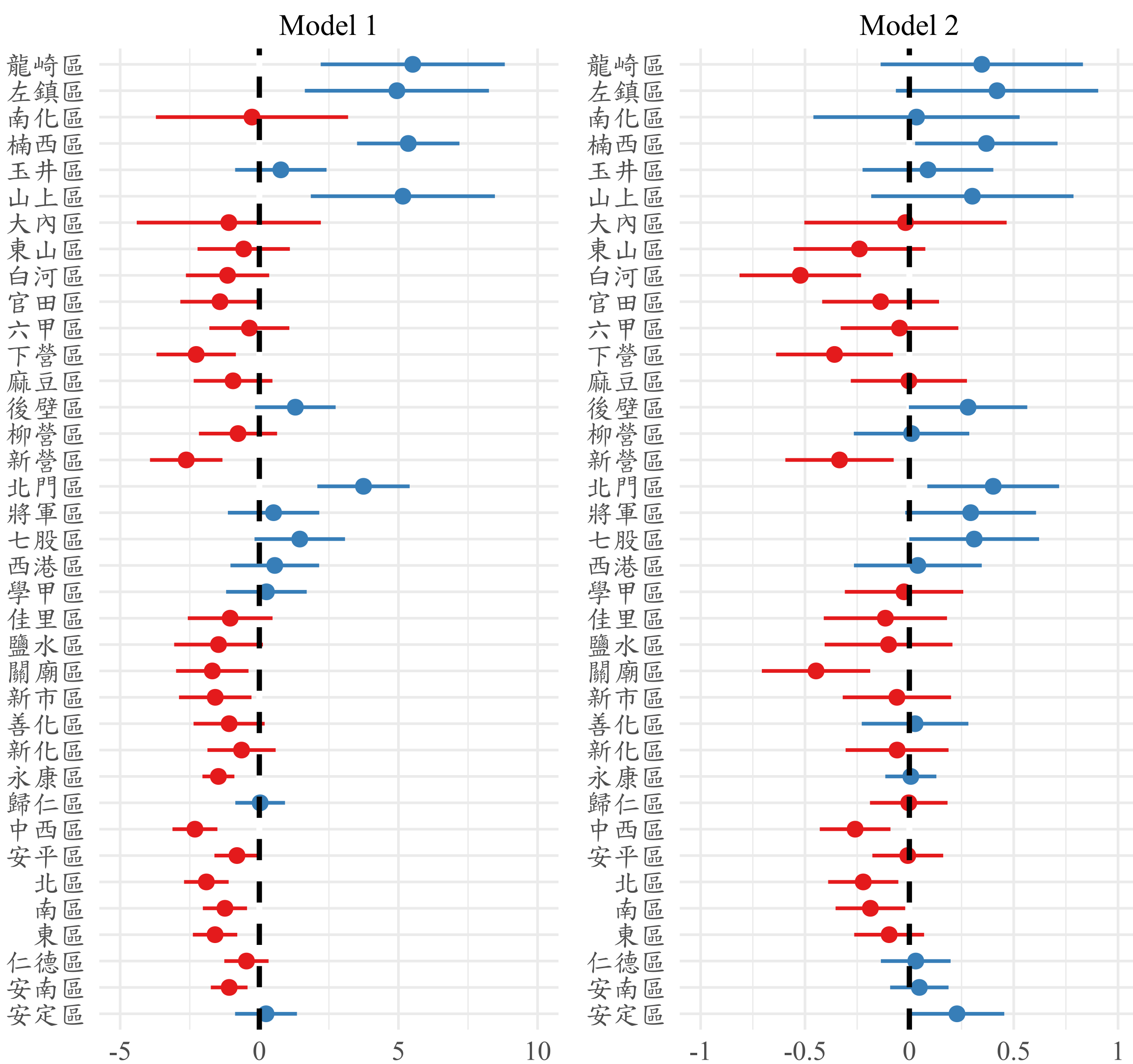

Figure 1: Random Effects of Intercept for each District

variation. It means the two variations are nearly the same when it approaches to 1, and it is no need to do multilevel regression. On the other hand, it indicates that the within-group Variation dominates when ICC approaches to 0. It is suggested to have ICC in range 0.05 to 0.25 for using multilevel model in the social science field. Under this background knowledge, ICC of Model 1 is 0.14, while 0.05 for Model 2, and thus, it is a better choice to keep going on Model 1.

Fixed-effect only models

The district variable is retrieved from the **database** managed by Department of Budget, Accounting and Statistics of Tainan City Government. The socioeconomic variables include population density, sex ratio, aging index, natural increase, and so forth. To develop the random intercept multilevel model, Model 3 chooses population density as the district-level variable, and the respondent variables contain driver license (owns or not) as well as the number of personal car (categorized into three types: no car, one car, and no less than two cars). Also, Model 4 uses population growth as the district-level variable instead, and adding the gender as well as age to be the respondent levels. And then, Model 5 adds the trip purpose variable, while further add the mode class in Model 6. The model results are shown in Table 2. The coefficient and the confidence interval of each fixed effect variable is shown in Figure 2.

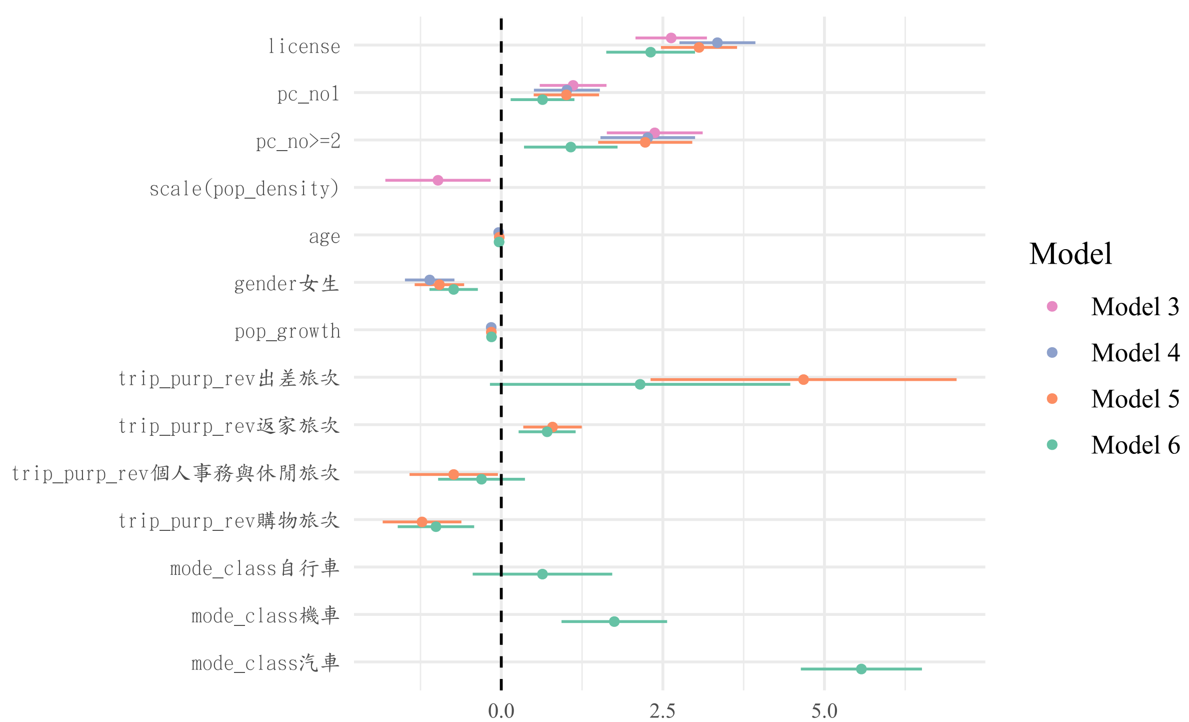

Figure 2: Coefficients of Fixed Effect Variable

For Model 3, we can find that respondents who have driver license would have 2.6 kilometers more on the trip distance. And the one who owns a personal car would have 1.1 kilometers more compared who have no cars. In terms of the coefficient of scale of population density, the negative (-0.979) value means that the trip distance would drop 0.979 kilometers as the population density increases by one standard deviation. It is because that the high population density area often owns a good accessibility to point interest, and people would travel less compared to the one living in a sparse area. Also, the intercept is 1.447, which is the fixed effect for all the district.

Table 2: Fixed-effect Only Models

	Model 3	Model 4	Model 5	Model 6
(Intercept)	1.447^{**}	2.619^{***}	2.354^{***}	1.292^{*}
	(0.506)	(0.552)	(0.546)	(0.614)
license	2.629^{***}	3.344^{***}	3.059^{***}	2.311^{***}
	(0.282)	(0.300)	(0.300)	(0.350)
pc_no1	1.112^{***}	1.015^{***}	1.007^{***}	0.638^{*}
	(0.263)	(0.260)	(0.258)	(0.252)
pc_no>=2	2.375^{***}	2.266^{***}	2.227^{***}	1.076^{**}
	(0.378)	(0.373)	(0.371)	(0.369)
scale(pop_density)	-0.979^{*}			
	(0.416)			
age		-0.040^{***}	-0.031^{***}	-0.034^{***}
		(0.006)	(0.006)	(0.006)
gender 女生		-1.108^{***}	-0.957^{***}	-0.737^{***}
		(0.195)	(0.195)	(0.191)
pop_growth		-0.156^{***}	-0.154^{***}	-0.152^{***}
		(0.037)	(0.036)	(0.035)
trip_purp_rev 出差旅次			4.677^{***}	2.148
			(1.208)	(1.185)
trip_purp_rev 返家旅次			0.793^{***}	0.709^{**}
			(0.231)	(0.225)
trip_purp_rev 個人事務與休閒旅次			-0.736^{*}	-0.306
			(0.349)	(0.343)
trip_purp_rev 購物旅次			-1.226^{***}	-1.011^{***}
			(0.311)	(0.302)
mode_class 自行車				0.637
				(0.551)
mode_class 機車				1.749^{***}
				(0.417)
mode_class 汽車				5.572^{***}
				(0.478)
AIC	21609.837	21532.349	21474.792	21270.243
BIC	21652.764	21587.540	21554.514	21368.362
Log Likelihood	-10797.918	-10757.174	-10724.396	-10619.122
Num. obs.	3403	3403	3403	3403
Num. groups: district	37	37	37	37
Var: district (Intercept)	3.617	3.246	3.057	2.830
Var: Residual	32.671	31.799	31.239	29.386

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

For Model 4, population growth is replaced with the population density (note that it may not appropriate to place in the same model, for they have a high correlation). We can find that the trip distance would decrease approximately 0.16 kilometers as the growth increases by one percentage. Age variable suggests that the older respondents, the shorter travel distance. Last, female has a relatively lower trip distance than the male concluded by the model result. As for Model 5, the trip purpose is further added. Note that for easily developing the model, we classify all the purpose into 5 types, including shopping, commuting, home-trip, leisure, and business. The base of the dummy variable is commuting trip, we can find that the trip distance of home-trip and business trip are significantly higher, while the leisure and shopping trip is relatively lower. Last, Model 6 adds a most vital variable: mode classification. We can find that the trip distance of scooter and car are all significantly higher.

Among all the 4 models mentioned above, AIC, BIC and log likelihood of Model 6 is apparently lowest. And it is mainly because the “mode” variable has been considered, and explains most of the variance.

Add Random Effects

In addition to the varying intercept model, varying slope supposes that the coefficient of each variable is different among districts. Trying to add random effects as much as possible, the final result is shown in Model 11 in 3. This model contains the license, number of personal car, and the scale of population growth introduced in the previous model. And the random effect variable includes whether the respondent is elderly (>65 years old), gender, and mode use in that trip. The diagnostics of the final model is shown in Figure 3.

From the diagnostics, we may conclude that the model performs not well enough, since the residuals are not very correspondent to the line in QQ plot, and the deviation is significant. In addition, the random effect of Model 11 is shown in 4.

The random effect of mode is obvious, which has a huge difference between districts. It may because that how mode usage influences the trip distance are totally different among districts. In 西港區, 佳里區, 白河區, the random effect of mode class is higher, which may imply that the trip distance is much more affected by the choice of modal in these districts. For instance, trip distance would increase a lot when using car, compared to 白河區 and 南化區 (where the random effects of car use is 0).

Table 3: Random-effect Models

	Model 7	Model 8	Model 9	Model 10	Model 11
(Intercept)	-0.359^{***}	-0.132	-0.164^{*}	0.485^{***}	0.607^{***}
	(0.089)	(0.081)	(0.083)	(0.080)	(0.077)
license	0.947^{***}	0.964^{***}	0.941^{***}	0.452^{***}	0.470^{***}
	(0.059)	(0.060)	(0.059)	(0.060)	(0.058)
pc_no1	0.237^{***}	0.277^{***}	0.239^{***}	0.160^{**}	0.135^{**}
	(0.055)	(0.056)	(0.055)	(0.051)	(0.050)
pc_no>=2	0.486^{***}	0.509^{***}	0.483^{***}	0.215^{**}	0.203^{**}
	(0.079)	(0.080)	(0.079)	(0.074)	(0.073)
scale(pop_growth)	-0.133^{**}	-0.061	-0.107^{**}	-0.076	-0.096^{**}
	(0.047)	(0.036)	(0.040)	(0.039)	(0.035)
AIC	10964.865	11066.216	10941.761	10488.042	10385.513
BIC	11020.057	11121.408	11015.350	10586.161	10551.088
Log Likelihood	−5473.433	−5524.108	−5458.880	−5228.021	−5165.756
Num. obs.	3403	3403	3403	3403	3403
Num. groups: district	37	37	37	37	37
Var: district (Intercept)	0.156	0.024	0.069	4.749	3.863
Var: district agedaged	0.492		0.474		0.299
Cov: district (Intercept) agedaged	−0.201		−0.121		1.036
Var: Residual	1.408	1.469	1.392	1.189	1.138
Var: district gender 女生		0.085	0.064		0.051
Cov: district (Intercept) gender 女生		0.045	0.019		0.266
Cov: district agedaged gender 女生			0.091		0.062
Var: district mode_class 自行車				1.826	1.665
Var: district mode_class 機車				4.064	3.445
Var: district mode_class 汽車				7.594	6.621
Cov: district (Intercept) mode_class 自行車				−2.926	−2.529
Cov: district (Intercept) mode_class 機車				−4.339	−3.605
Cov: district (Intercept) mode_class 汽車				−5.961	−5.001
Cov: district mode_class 自行車 mode_class 機車				2.708	2.387
Cov: district mode_class 自行車 mode_class 汽車				3.638	3.247
Cov: district mode_class 機車 mode_class 汽車				5.450	4.619
Cov: district agedaged mode_class 自行車					−0.669
Cov: district agedaged mode_class 機車					−0.952
Cov: district agedaged mode_class 汽車					−1.397
Cov: district gender 女生 mode_class 自行車					−0.165
Cov: district gender 女生 mode_class 機車					−0.209
Cov: district gender 女生 mode_class 汽車					−0.327

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

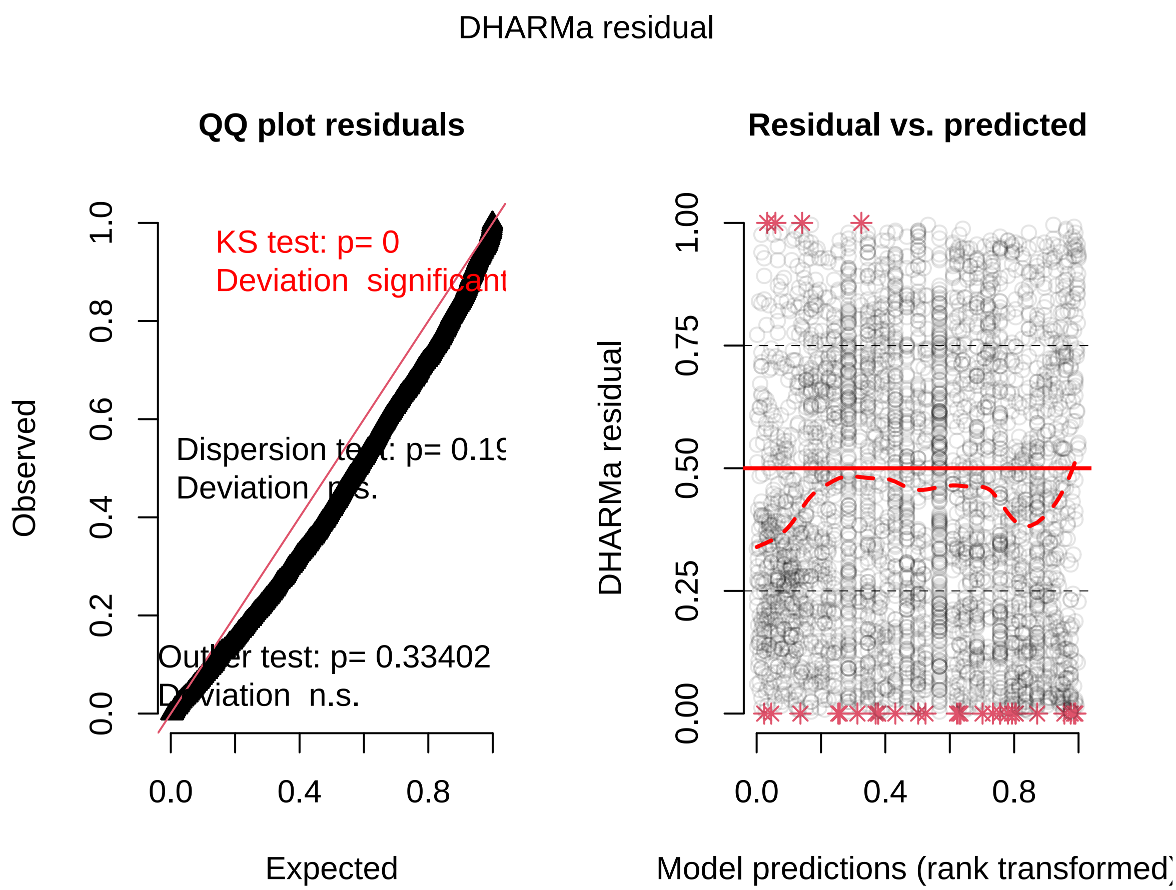

Figure 3: Diagnostics of Model 11

Random effects

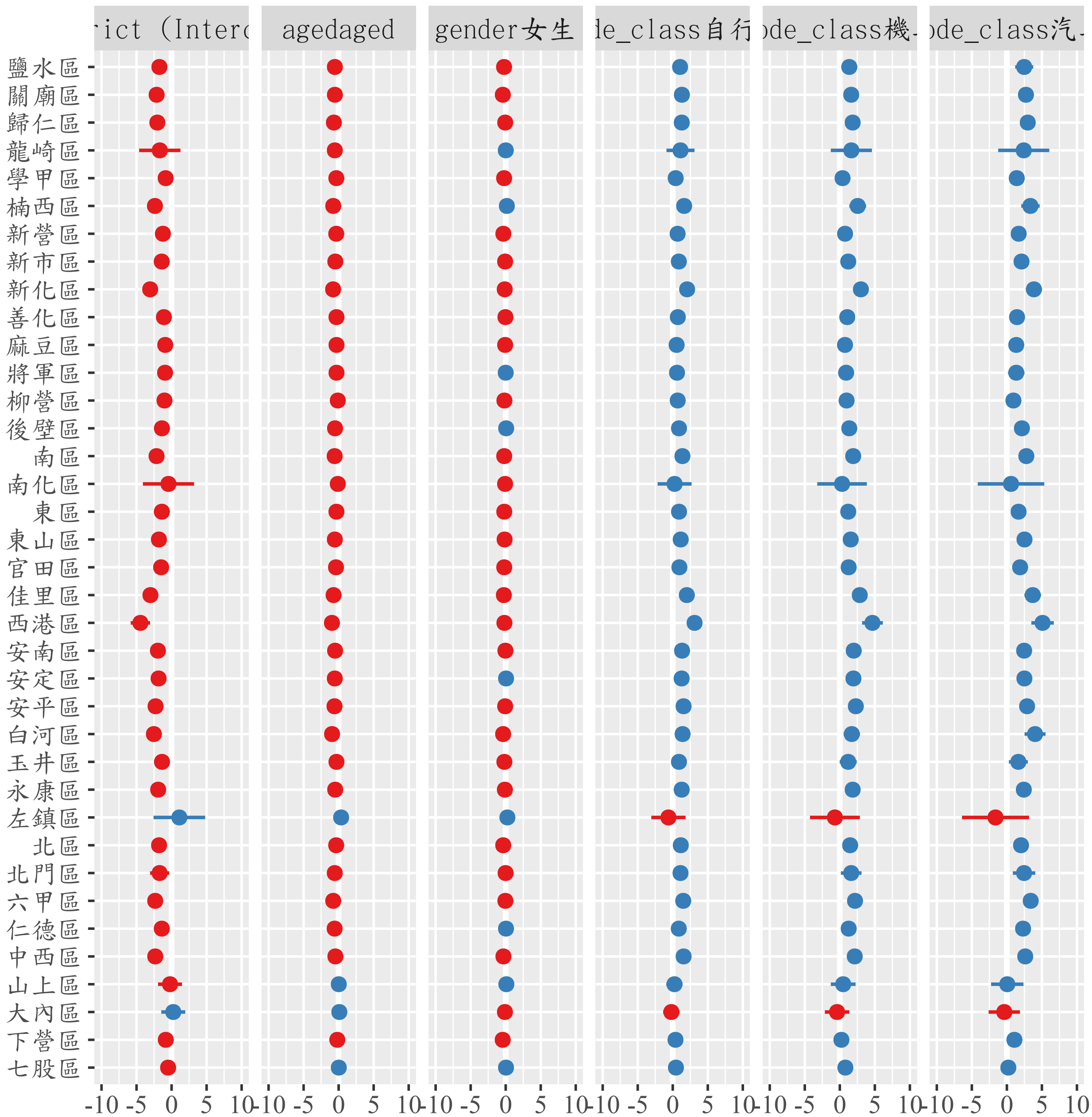

Figure 4: Diagnostics of Model 11