

Supplement 4: sensitivity analysis for Pätši et al.

Arho Toikka

2023-03-14

Assumption checks & sensitivity analyses

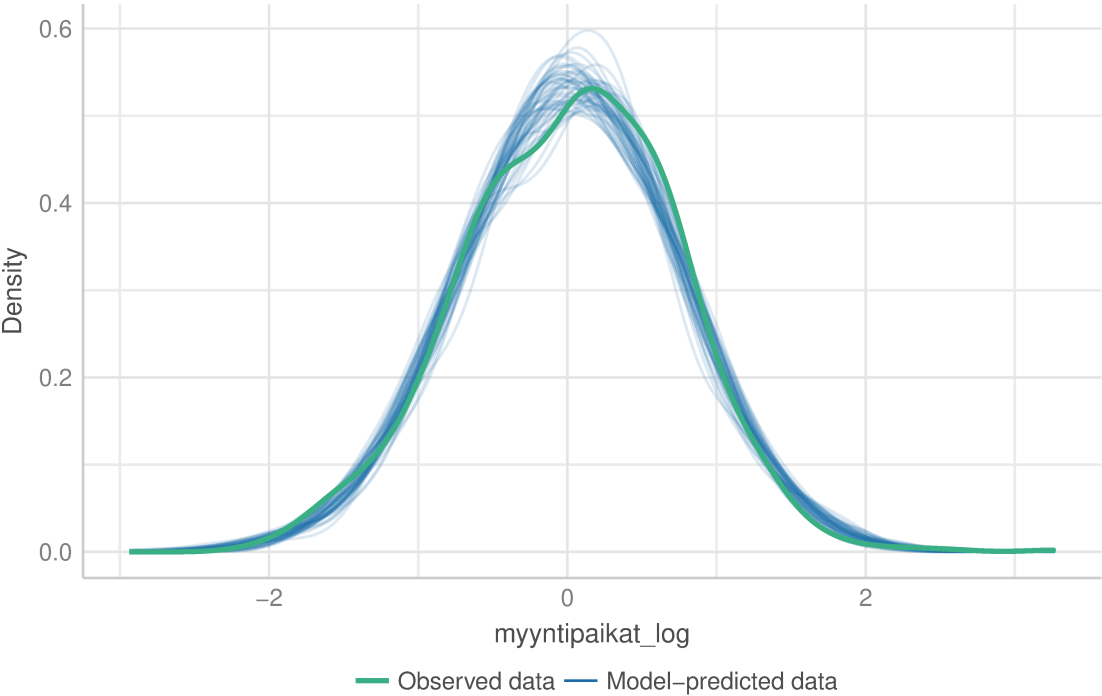
Posterior predictive checks are within expected range.

Linearity, homogeneity of variance and leverage identify a potential problem data point. Linearity also shows a minor uptick at the low end. To see the effect of the influential data point, Sensitivity analysis #1: removes outliers.

Median income and % of low income are potentially multicollinear. To see the effect of multicollinearity, Sensitivity analysis #2: use only one income variable.

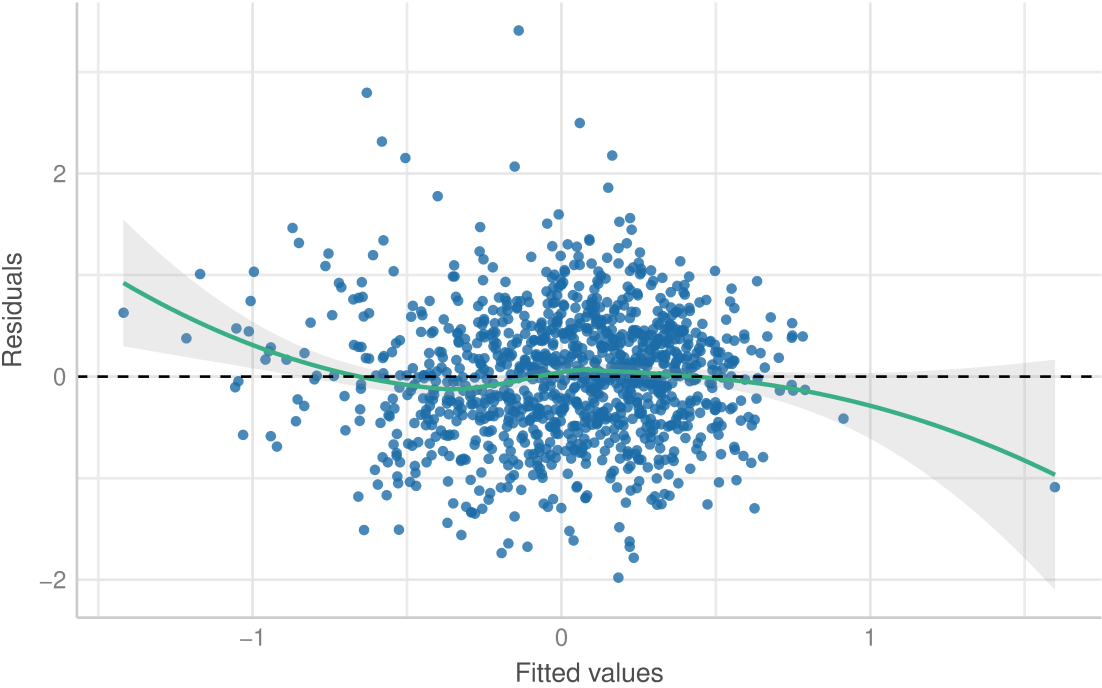
\$PP_CHECK

Posterior Predictive Check
Model-predicted lines should resemble observed data line



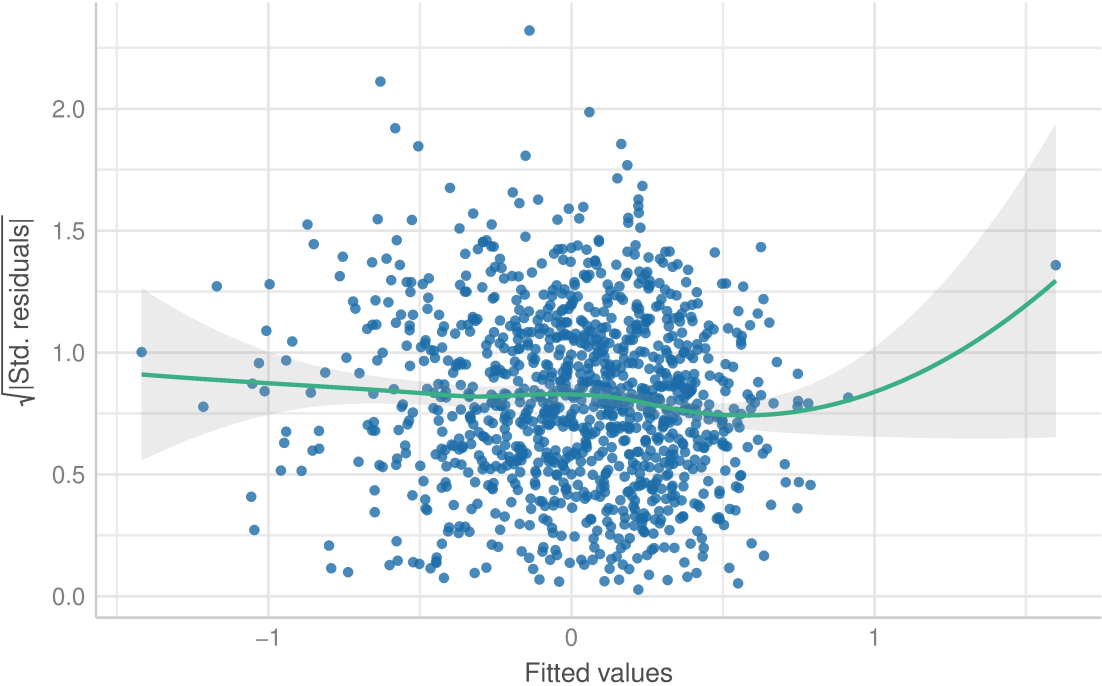
\$NCV

Linearity
Reference line should be flat and horizontal



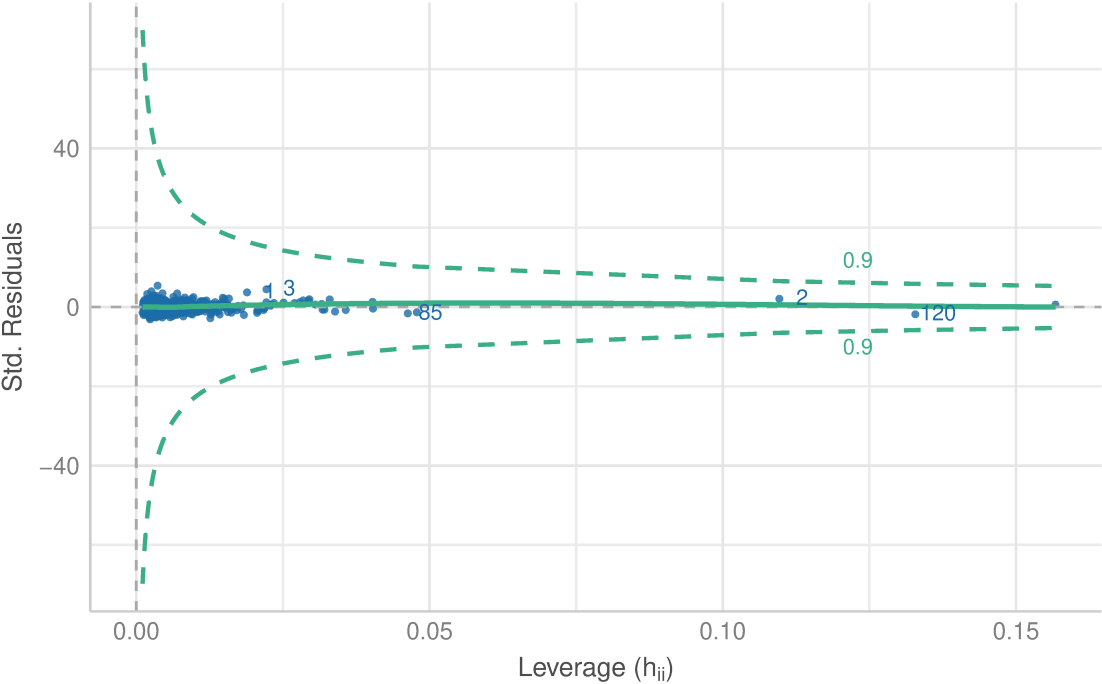
\$HOMOGENEITY

Homogeneity of Variance
Reference line should be flat and horizontal



\$OUTLIERS

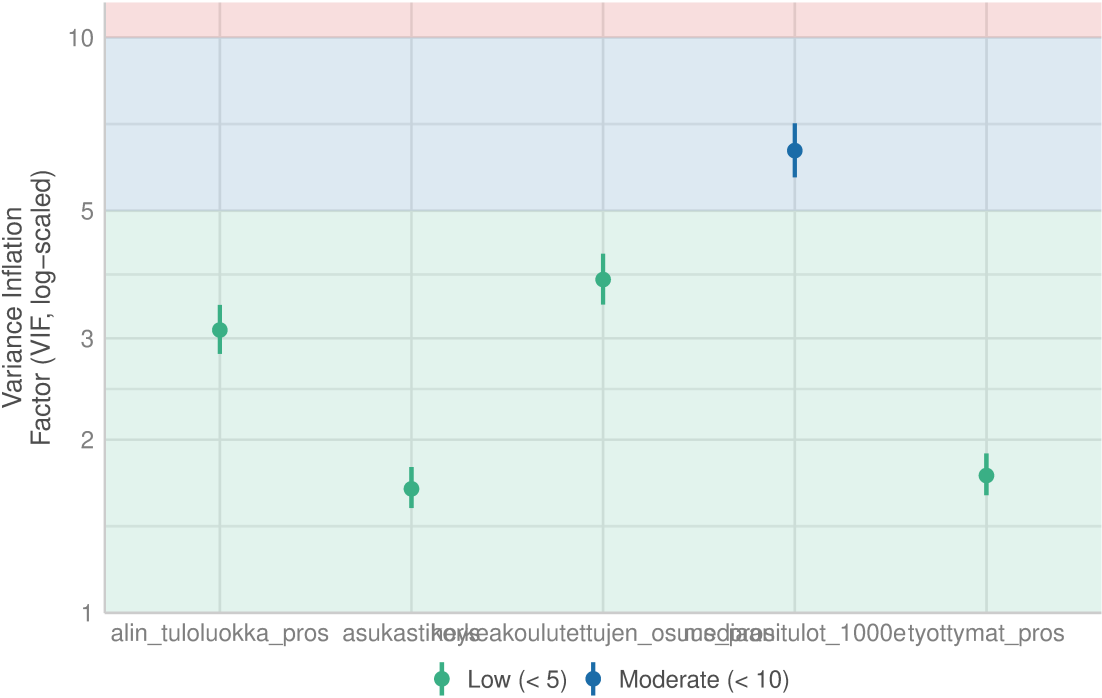
Influential Observations
Points should be inside the contour lines



\$VIF

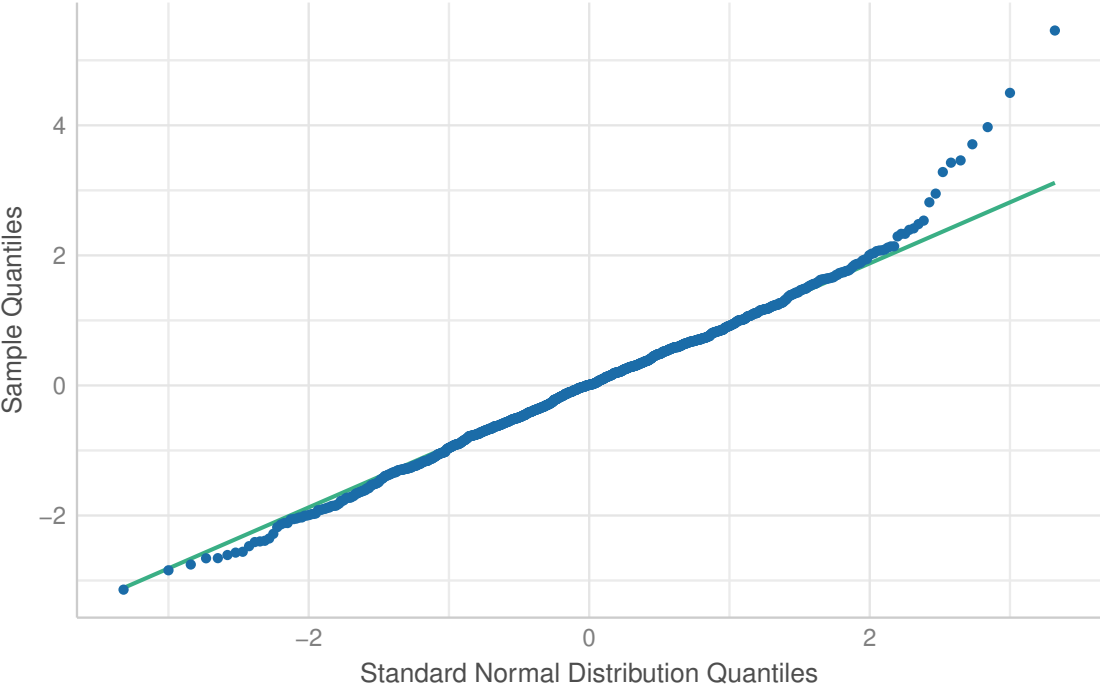
Collinearity

High collinearity (VIF) may inflate parameter uncertainty



\$QQ

Normality of Residuals
Dots should fall along the line



Sensitivity #1: Remove high leverage points

The influential data points are those with an unusually high density of retail locations. Removing all locations with $\log(\text{density}) > 2.5$ or density > 12.18 to see if this changes results.

Coefficient changes are small and removal of outliers/high leverage points does not affect conclusions.

Table 1: Multivariate associations for presence of retailer and retailer density by SES, density outliers removed

| Variable | Logistic regression for presence of retailer | | | Linear regression for log of retailer density | | |
|---|--|-------|---------|---|-------|---------|
| | Estimate | SE | P value | Estimate | SE | P value |
| Median income, 1000 euro | 0.543 | 0.060 | <0.001 | 0.957 | 0.014 | 0.002 |
| % in lowest income category | 0.885 | 0.028 | <0.001 | 1.029 | 0.007 | <0.001 |
| % unemployed | 1.064 | 0.026 | 0.018 | 0.977 | 0.006 | <0.001 |
| % with higher education | 1.139 | 0.017 | <0.001 | 0.983 | 0.003 | <0.001 |
| Population density, 1000s per square km | 2.911 | 0.250 | <0.001 | 1.033 | 0.014 | 0.020 |

Estimates are exponentiated and can be interpreted as multiplicative increases in probability and density, given 1 unit change in predictor

Sensitivity #2: remove potentially collinear variables

Conventional guidelines for VIF are that if the largest VIF is >10, there is evidence of consequential multicollinearity, and if >5 there is cause for concern. However, it is important to note that as O’Brien (2007) shows VIF values as high as 40 do not necessarily mean the regression is wrong or call for the elimination of one or more independent variables, and best practice is to expli

Here are the VIF values. For both models, they are lower than 10, but for income they are higher than 5, meaning there is some reason for concern. Sensitivity analysis #2 above explores what happens if we remove the highest VIF variable. For the linear model, removing one of the income variables does not change the results in a way that would change the conclusions. For the logistic model, removing median income changes the direction of the lowest income category from negative and significant to positive and non-significant. This has little substantial effect on the conclusions.

| | |
|--------------------------------|----------|
| | x |
| mediaanitulot_1000e | 6.358342 |
| alin_tuloluokka_pros | 3.099495 |
| tyottomat_pros | 1.731645 |
| korkeakoulutettujen_osuus_pros | 3.796562 |
| asukastiheys | 1.642141 |

| | |
|--------------------------------|----------|
| | x |
| mediaanitulot_1000e | 6.869882 |
| alin_tuloluokka_pros | 2.922583 |
| tyottomat_pros | 1.702085 |
| korkeakoulutettujen_osuus_pros | 4.483845 |
| asukastiheys | 1.475021 |

| | Logistic regression for presence of retailer | | | Linear regression for log of retailer density | | |
|---|--|-------|---------|---|-------|---------|
| Variable | Estimate | SE | P value | Estimate | SE | P value |
| % in lowest income category | 1.047 | 0.022 | 0.038 | 1.043 | 0.005 | <0.001 |
| % unemployed | 1.178 | 0.025 | <0.001 | 0.981 | 0.006 | <0.001 |
| % with higher education | 1.023 | 0.012 | 0.060 | 0.976 | 0.003 | <0.001 |
| Population density, 1000s per square km | 3.092 | 0.229 | <0.001 | 1.041 | 0.014 | 0.005 |

\caption{Multivariate associations for presence of retailer and retailer density by SES, % in lowest income
dropped}

| | Logistic regression for presence of retailer | | | Linear regression for log of retailer density | | |
|---|--|-------|---------|---|-------|---------|
| Variable | Estimate | SE | P value | Estimate | SE | P value |
| Median income, 1000€ | 0.620 | 0.050 | <0.001 | 0.917 | 0.010 | <0.001 |
| % unemployed | 1.056 | 0.026 | 0.039 | 0.977 | 0.006 | <0.001 |
| % with higher education | 1.137 | 0.017 | <0.001 | 0.986 | 0.003 | <0.001 |
| Population density, 1000s per square km | 2.843 | 0.245 | <0.001 | 1.033 | 0.014 | 0.026 |

\end{table}

7

population density are brought closer to 1 or no effect and become non-significant, but the others remain similar and significant.

Table 3: Multivariate associations for presence of retailer and retailer density by SES, general linear model with log-link

| Variable | Logistic regression for presence of retailer | | | General linear model with log-link | | |
|---|--|-------|---------|------------------------------------|-------|---------|
| | Estimate | SE | P value | Estimate | SE | P value |
| Median income, 1000s € | 0.542 | 0.060 | <0.001 | 0.933 | 0.026 | 0.007 |
| % in lowest income category | 0.884 | 0.028 | <0.001 | 1.004 | 0.010 | 0.696 |
| % unemployed | 1.063 | 0.026 | 0.020 | 0.979 | 0.008 | 0.007 |
| % with higher education | 1.139 | 0.017 | <0.001 | 0.987 | 0.005 | 0.018 |
| Population density, 1000s per square km | 2.903 | 0.250 | <0.001 | 0.993 | 0.030 | 0.809 |

Estimates are exponentiated and can be interpreted as multiplicative increases in probability and density, given 1 unit change in predictor

Sensitivity check #4: Zero-inflated Gamma-regression

Another alternative model would be zero-inflated Gamma-regression, a model that models both for the absence of retailer (note: this is the opposite of the logistic model, which models for presence of retailer) and the density at the same time, in one model.

Results are the same as the models in the main paper (to compare the zero-inflated coefficients to the logistic regression coefficients, take the reciprocal 1/coefficient).

Table 4: Multivariate associations for presence of retailer and retailer density by SES, zero-inflated gamma

| component | term | estimate | std.error | statistic | p.value |
|-----------|---|----------|-----------|-----------|---------|
| cond | Intercept | 2.781 | 0.389 | 2.627 | 0.009 |
| cond | Median income, 1000s € | 0.964 | 0.014 | -2.599 | 0.009 |
| cond | % in lowest income category | 1.033 | 0.007 | 4.699 | 0.000 |
| cond | % unemployed | 0.985 | 0.004 | -4.111 | 0.000 |
| cond | % with higher education | 0.970 | 0.006 | -4.868 | 0.000 |
| cond | Population density, 1000s per square km | 1.035 | 0.014 | 2.495 | 0.013 |
| zi | Intercept | 0.000 | 1.670 | -8.631 | 0.000 |
| zi | Median income, 1000s € | 1.845 | 0.060 | 10.244 | 0.000 |
| zi | % in lowest income category | 1.132 | 0.028 | 4.423 | 0.000 |
| zi | % unemployed | 0.878 | 0.017 | -7.561 | 0.000 |
| zi | % with higher education | 0.941 | 0.026 | -2.335 | 0.020 |
| zi | Population density, 1000s per square km | 0.345 | 0.250 | -4.258 | 0.000 |

^a cond = density, zi = Probability for NO sales

Sensitivity check #5: three income tertiles

To see if including % of high income along with % of low income, essentially making middle income a comparison group, that was tested. However, the high correlations between high income % and low income % make variance inflation issues worse and add instability to estimates.

“Median income, 1000€” = mediaanitulos_1000e % in lowest income category = alin_tuloluokka_pros “% unemployed” = tyottomat_pros “% with higher education” = korkeakoulutettujen_osuus_pros
 “Population density, 1000s per square km” = asukastiheys

| term | estimate | std.error | statistic | p.value |
|----------------------|------------|-----------|-----------|---------|
| (Intercept) | -1.1261242 | 0.0838618 | -13.42833 | 0 |
| alin_tuloluokka_pros | 0.0574723 | 0.0040850 | 14.06899 | 0 |

| term | estimate | std.error | statistic | p.value |
|----------------------|------------|-----------|-----------|-----------|
| (Intercept) | -0.2473787 | 0.1499824 | -1.649385 | 0.0993523 |
| alin_tuloluokka_pros | 0.0332116 | 0.0052918 | 6.276094 | 0.0000000 |
| high_income | -0.0208686 | 0.0029807 | -7.001174 | 0.0000000 |

| term | estimate | std.error | statistic | p.value |
|----------------------|------------|-----------|------------|-----------|
| (Intercept) | -0.1099712 | 0.2836364 | -0.3877188 | 0.6982241 |
| alin_tuloluokka_pros | 0.0689464 | 0.0147030 | 4.6892784 | 0.0000027 |

| term | estimate | std.error | statistic | p.value |
|----------------------|------------|-----------|-----------|-----------|
| (Intercept) | 1.4806240 | 0.5124606 | 2.889245 | 0.0038617 |
| alin_tuloluokka_pros | 0.0225520 | 0.0188901 | 1.193853 | 0.2325356 |
| high_income | -0.0341776 | 0.0092772 | -3.684040 | 0.0002296 |

| term | estimate | std.error | statistic | p.value |
|--------------------------------|------------|-----------|------------|-----------|
| (Intercept) | 2.2323313 | 0.7157267 | 3.1189716 | 0.0018618 |
| mediaanitulos_1000e | -0.0694592 | 0.0255623 | -2.7172494 | 0.0066856 |
| alin_tuloluokka_pros | 0.0038414 | 0.0098272 | 0.3908955 | 0.6959498 |
| tyottomat_pros | -0.0213044 | 0.0078538 | -2.7126108 | 0.0067792 |
| korkeakoulutettujen_osuus_pros | -0.0129035 | 0.0054688 | -2.3594901 | 0.0184732 |
| asukastiheys | -0.0072266 | 0.0299004 | -0.2416895 | 0.8090656 |

| term | estimate | std.error | statistic | p.value |
|--------------------------------|------------|-----------|-----------|-----------|
| (Intercept) | 3.4471968 | 0.6592186 | 5.229217 | 0.0000002 |
| mediaanitulos_1000e | -0.1660028 | 0.0301337 | -5.508874 | 0.0000000 |
| alin_tuloluokka_pros | 0.0093624 | 0.0078749 | 1.188888 | 0.2347394 |
| high_income | 0.0451081 | 0.0098383 | 4.584962 | 0.0000051 |
| tyottomat_pros | -0.0232745 | 0.0059939 | -3.883050 | 0.0001093 |
| korkeakoulutettujen_osuus_pros | -0.0231226 | 0.0035690 | -6.478758 | 0.0000000 |
| asukastiheys | 0.0322759 | 0.0142543 | 2.264291 | 0.0237495 |

| | X |
|--------------------------------|-----------|
| mediaanitulos_1000e | 27.443896 |
| alin_tuloluokka_pros | 4.159069 |
| high_income | 20.459588 |
| tyottomat_pros | 1.737190 |
| korkeakoulutettujen_osuus_pros | 4.244679 |
| asukastiheys | 1.642769 |

| term | estimate | std.error | statistic | p.value |
|--------------------------------|------------|-----------|------------|-----------|
| (Intercept) | 14.4118729 | 1.6695759 | 8.632056 | 0.0000000 |
| mediaanitulos_1000e | -0.6124198 | 0.0597751 | -10.245405 | 0.0000000 |
| alin_tuloluokka_pros | -0.1236361 | 0.0279513 | -4.423275 | 0.0000097 |
| tyottomat_pros | 0.0612427 | 0.0262303 | 2.334810 | 0.0195533 |
| korkeakoulutettujen_osuus_pros | 0.1298752 | 0.0171756 | 7.561591 | 0.0000000 |
| asukastiheys | 1.0656341 | 0.2502250 | 4.258703 | 0.0000206 |

| term | estimate | std.error | statistic | p.value |
|--------------------------------|------------|-----------|------------|-----------|
| (Intercept) | 15.8387385 | 2.4103911 | 6.5710244 | 0.0000000 |
| mediaanitulos_1000e | -0.6846577 | 0.1062445 | -6.4441678 | 0.0000000 |
| alin_tuloluokka_pros | -0.1352138 | 0.0313498 | -4.3130694 | 0.0000161 |
| high_income | 0.0271425 | 0.0328255 | 0.8268727 | 0.4083092 |
| tyottomat_pros | 0.0622116 | 0.0262678 | 2.3683622 | 0.0178670 |
| korkeakoulutettujen_osuus_pros | 0.1259852 | 0.0177514 | 7.0971925 | 0.0000000 |
| asukastiheys | 1.0639583 | 0.2500560 | 4.2548798 | 0.0000209 |

| | X |
|--------------------------------|-----------|
| mediaanitulot_1000e | 21.591972 |
| alin_tuloluokka_pros | 3.665652 |
| high_income | 14.850653 |
| tyottomat_pros | 1.705356 |
| korkeakoulutettujen_osuus_pros | 4.774928 |
| asukastiheys | 1.470565 |