**BIOST 2050: Longitudinal and Clustered Data Analysis**

**Homework Assignment #1 (part 2)**

**Due: Friday, September 22, 2023**

Answer the questions and justify your answers. Note that you will receive a major deduction if you answer a question by giving only the software output without justifying your answer.

Enterprise zones (EZs) are geographic areas in which companies can qualify for a variety of incentives, subsidies, or tax credits. EZ programs are established to encourage businesses to remain, locate, or expand in areas with high unemployment and significant poverty, thereby aiding in the revitalization of these depressed regions.

A study was conducted to analyze panel data concerning an enterprise zone program in the state of Indiana. Data were gathered from 22 area unemployment claims offices from 1980 to 1988. Of these, six were designated as EZs starting in 1984 and an additional four began their EZ designation in 1985. The remaining 12 unemployment claims offices did not have an EZ in their area during the study's timeframe.

A region was selected as an EZ based on several factors: the population and geographic size of the area, the economic indicators (such as unemployment and poverty rates), and the decision of the state government. An EZ designation is typically intended to last 10 years.

Analytic dataset enterprise.dta contains the following variables:

**Variable Label**

area: unemployment claims area identifier ()

year: calendar year ()

uclms: number of unemployment claims ()

uclms1: baseline number of unemployment claims ()

ez: designation as an enterprise zones (0=no, 1=yes) ()

time:time 1, …, 9 () corresponding to calendar years 1980-1988

1. Let be the number of unemployment claims for area () at calendar time (). Fit a model to investigate the first study aim: **relationship between the number of unemployment claims and calendar year; and how the relationship varies when comparing areas included in the enterprise zones with those that are not included in an enterprise zone.**
   1. Write down the formula of the model assuming random area-specific intercepts. Explicitly specify the fixed effects and the random effects, if any, and justify your choices. (12 pts)

uclmsij – uclms1j = beta\_0 + zeta\_j + beta\_1\*time + beta\_2\*ez + beta\_3\*ez\*time + episilonij

where zeta\_j = random-intercept, and betas are fixed intercepts/effects

* 1. **Fit the model in 3(a) using a statistical package. (8 pts)**

> ri\_fit\_mle <- lmerTest::lmer(uclms\_diff ~ 1 + time\*ez + (1 | area), data = enterprise, REML = FALSE)

> summary(ri\_fit\_mle)

Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's method ['lmerModLmerTest']

Formula: uclms\_diff ~ 1 + time \* ez + (1 | area)

Data: enterprise

AIC BIC logLik deviance df.resid

4803.4 4823.2 -2395.7 4791.4 192

Scaled residuals:

Min 1Q Median 3Q Max

-4.0392 -0.5510 -0.0138 0.4257 5.8937

Random effects:

Groups Name Variance Std.Dev.

area (Intercept) 1.114e+09 33381

Residual 1.510e+09 38856

Number of obs: 198, groups: area, 22

Fixed effects:

Estimate Std. Error df t value Pr(>|t|)

(Intercept) 17394.31 9568.08 51.64 1.818 0.0749 .

time -12025.06 1354.70 187.13 -8.877 5.56e-16 \*\*\*

ez -32478.06 32214.91 177.59 -1.008 0.3147

time:ez 3235.56 4507.24 177.48 0.718 0.4738

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

(Intr) time ez

time -0.571

ez -0.120 0.123

time:ez 0.172 -0.297 -0.955

* 1. **Based on the fitted model in 3(b), is there a time effect? Describe it and specify the null and alternative hypotheses associated with the test. Note: please describe the time effect thoroughly (that is, interpret your finding and not just simply say “there is a (or there is no) significant time effect based on the p value <0.05 (or ≥0.05).” (6 pts)**

Yes, there is a significant time effect. Assuming panel data, each increment in time (1 year) isis associated with a 12,025 person decrease in the expected difference of unemployment claims relative to baseline unemployment claims.

* 1. **Does the relationship between the number of unemployment claims and time vary when comparing areas included in the enterprise zones with those that are not included in an enterprise zone? Answer this question by specifying the null and alternative hypotheses and the associated test results. (6 pts)**

The null hypothesis is that the fixed effect (beta\_3) coefficient of the time by enterprise zone interaction is zero. The alternative hypothesis is that this coefficient is not equal to zero. This can be tested with a t-test, where t = beta\_3\_estimate / standard\_error(beta\_3\_estimate) given in the output. This statistical test does not reject the null hypothesis in my model.

* 1. **Based on the fitted model in 3(b), how much of the total variation in the number of unemployment claims is attributable to area variability? (6 pts)**

ri\_fit\_reml <- lmerTest::lmer(uclms\_diff ~ 1 + time\*ez + (1 | area), data = enterprise)

summary(ri\_fit\_reml)

performance::icc(ri\_fit\_reml)

# Intraclass Correlation Coefficient

Adjusted ICC: 0.434

Unadjusted ICC: 0.312

The proportion of total variance in change in number of unemployment claims due to inter-area variation is 43.4%

1. Fit a model to investigate the second study aim: **whether the inclusion in the enterprise zones would reduce the number of unemployment claims.**
   1. Write down the formula of the model assuming random area-specific intercepts, random area-specific time effects, and possible intercept-slope covariance. (12 pts)

uclmsij – uclms1j = beta\_0 + zeta\_0j + beta\_1\*timeij + zeta\_1j\*time\_ij + beta\_2\*ezij + beta\_3\*ezij\*time + episilonij

where zeta\_0j = random-intercept for area

where zeta\_1j = random area-specific time effect

* 1. **Fit the model using a statistical package. (8 pts)**

> rc\_fit\_mle <- lmerTest::lmer(uclms\_diff ~ 1 + time\*ez + (1 + time || area), data = enterprise, REML = FALSE)

> summary(rc\_fit\_mle)

Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's method ['lmerModLmerTest']

Formula: uclms\_diff ~ 1 + time \* ez + (1 + time || area)

Data: enterprise

AIC BIC logLik deviance df.resid

4761.5 4784.5 -2373.7 4747.5 191

Scaled residuals:

Min 1Q Median 3Q Max

-1.7673 -0.3899 -0.0312 0.2883 6.2220

Random effects:

Groups Name Variance Std.Dev.

area (Intercept) 1.892e+08 13754

area.1 time 4.636e+07 6809

Residual 1.096e+09 33103

Number of obs: 198, groups: area, 22

Fixed effects:

Estimate Std. Error df t value Pr(>|t|)

(Intercept) 15239.35 6509.20 44.31 2.341 0.0238 \*

time -11119.33 2043.41 45.81 -5.442 1.99e-06 \*\*\*

ez -27597.75 27564.70 165.32 -1.001 0.3182

time:ez 1131.08 4267.55 191.16 0.265 0.7913

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

(Intr) time ez

time -0.535

ez -0.192 0.128

time:ez 0.331 -0.337 -0.908

* 1. **For a specific calendar year, estimate the conditional covariance of the number of unemployment claims between two different areas and . (6 pts)**

The unemployment claims in two different areas, given their covariates, are independent for any given calendar year. Therefore, their conditional covariance is zero.

* 1. **Based on the fitted model in 4(b), would inclusion in the enterprise zones reduce the number of unemployment claims? Answer this question by specifying the null and alternative hypotheses associated and the associated test results. (6 pts)**

The null hypothesis is that the fixed effect (beta\_2) coefficient of enterprise zones is zero (i.e., there is no effect of enterprise zones on expected difference in unemployment claims). The alternative hypothesis is that this coefficient is not equal to zero. This can be tested with a t-test, where t = beta\_2\_estimate / standard\_error(beta\_2\_estimate) given in the model summary in b). This statistical test does not reject the null hypothesis in my model. This result holds regardless of whether the interaction term between time and ez is included or excluded in the model.