Longitudinal Homework 6

Tim Vigers

12 November 2019

1. Model planning

a. Software

i. Medication use

In order to adapt a generalized linear model (GzLM) for serially correlated data, I would use a generalized estimating equation (GEE). This could be fit using PROC GENMOD in SAS, with a Poisson distribution since this is count data. There is also a package called "geepack" in R that can fit GEEs, but I've never used it and don't know much about it, so PROC GENMOD is probably safer.

ii. FEV1

I think you can probably get away with a normal theory model for FEV1 data, unless it's really skewed. If so, I would use either PROC MIXED in SAS or gls() in R. If normal theory models won't work, then I would use PROC GENMOD or geeglm() like above, to model the outcome with a non-normal distribution.

b. Data

i. Medication use

Because we need a GEE for this outcome, I would set up the data so that each subject has a row for every day during the relevant timeframe. On days without an albuterol count the outcome would be filled in as missing (NA in R), but the temporal spacing would be equal between rows within subject.

The SAS code would look something like:

```
proc genmod data=albuterol;
class id friday;
model albuterol_use =
  friday ln_mmax_pm25 temperature pressure humidity / solution dist=poisson;
repeated subject=id / TYPE = AR(1);
run;
```

c. Binary outcome

i. Medication use

In order to fit a GzLMM model with a random intercept for subject, I would use PROC GLIMMIX in SAS (with distribution = binary). PROC NLMIXED would also work, but I find it a little more confusing. The default in PROC GLIMMIX is to approximate the true likelihood using Laplace's method, but you can also specify method = quad to use Gaussian quadrature. This can also be done using glmer() in R.

ii. FEV1

Again assuming that the normal theory model is okay for FEV1, I would use PROC MIXED in SAS or lme() in R (or possible lmer() to be consistent with the model above). These models would be fit using ML or REML rather than Gaussian quadrature.

for an AR(1) correlation structure in the repeated measures, but does not allow for the incorporation of random effects. GEEs can be fit (using the same data structure as above) with:

##d. Random intercept and serial correlation

i. Medication use

Correlated count data like this probably requires a generalized linear mixed model (GzLMM) where the outcome is modeled as Poisson-distributed, a random intercept for subject, and with an AR(1) or spatial power correlation for repeated measures. I think the best way to do this in SAS is to use PROC GLIMMIX, and in R glmmPQL() should work.

There isn't a REPEATED statement in PROC GLIMMIX, so you need to another random effect with the "*residual*" and the correlation structure:

```
proc glimmix data=albuterol;
model albuterol_use =
  friday ln_mmax_pm25 temperature pressure humidity / solution distribution=poisson;
random intercept / subject=id;
random _residual_ / subject=id type=ar(1);
run;
```

The R code would be something like:

```
glmmPQL(fixed = albuterol_use ~ friday + ln_mmax_pm25 + temperature + pressure + humidity,
        random = ~1|id, family = "poisson",correlation = corAR1(),
        data = df)
```

ii. FEV1

These normal models could be fit using either PROC MIXED in SAS or lme() in R. I don't know how to include the AR(1) structure in lmer(), but I'm sure it's possible.

The SAS code would be something like:

```
proc mixed data=albuterol;
class id friday;
model fev1 =
  friday ln_mmax_pm25 temperature pressure humidity / solution;
random intercept / subject=id;
repeated / type=AR(1) subject=id;
run;
```

And the R code something like:

```
lme(fev1 ~ friday + ln_mmax_pm25 + temperature + pressure + humidity,
    random = ~1|id,correlation = corAR1(),data = df,na.action = na.omit)
```

2. Albuterol data

a. GEE

The results of PROC GENMOD are below, with the scale parameter in the red box:

Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		-8.3451	3.4749	-15.1559	-1.5344	-2.40	0.0163
temperature		-0.0063	0.0013	-0.0089	-0.0037	-4.69	<.0001
pressure		0.0013	0.0019	-0.0024	0.0051	0.69	0.4886
humidity		-0.0039	0.0006	-0.0052	-0.0027	-6.17	<.0001
friday	0	1.2167	0.0810	1.0579	1.3754	15.02	<.0001
friday	1	0.0000	0.0000	0.0000	0.0000	.	.
date		0.0004	0.0002	0.0000	0.0008	2.09	0.0368
ln_mmax_pm25		0.0532	0.0151	0.0237	0.0828	3.53	0.0004

Analysis Of GEE Parameter Estimates							
Model-Based Standard Error Estimates							
Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		-8.3451	4.5588	-17.2801	0.5899	-1.83	0.0672
temperature		-0.0063	0.0015	-0.0093	-0.0033	-4.16	<.0001
pressure		0.0013	0.0026	-0.0038	0.0065	0.51	0.6132
humidity		-0.0039	0.0009	-0.0056	-0.0022	-4.58	<.0001
friday	0	1.2167	0.0376	1.1429	1.2904	32.34	<.0001
friday	1	0.0000	0.0000	0.0000	0.0000	.	.
date		0.0004	0.0003	-0.0001	0.0009	1.75	0.0802
ln_mmax_pm25		0.0532	0.0193	0.0155	0.0910	2.76	0.0057
Scale		0.8573	.	.	.	.	.

Adding the scale parameter generally increases the standard errors (except for friday = 0).

b. GzLMM

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	id	0.5458	0.1135
SP(EXP)	id	0.6037	0.03412
Residual		0.6630	0.01272

Solutions for Fixed Effects						
Effect	friday	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		-0.6693	1.4582	56	-0.46	0.6480
friday	0	1.2028	0.03845	5915	31.29	<.0001
friday	1	0	.	.	.	.
ln_mmax_pm25		0.03787	0.01901	5915	1.99	0.0464
temperature		-0.00479	0.001176	5915	-4.07	<.0001
pressure		-0.00032	0.002311	5915	-0.14	0.8889
humidity		-0.00263	0.000738	5915	-3.57	0.0004

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
friday	1	5915	978.80	<.0001
ln_mmax_pm25	1	5915	3.97	0.0464
temperature	1	5915	16.56	<.0001
pressure	1	5915	0.02	0.8889
humidity	1	5915	12.71	0.0004

The residual estimate is equivalent to the square of the scale parameter, so $\sqrt{0.6630} = 0.814$. This is a little bit lower than the scale parameter from the GEE (0.857), but I'd say they're pretty close.

c. Dispersion

These scale estimates suggest underdispersion because they are less than 1.

d. Slopes and SEs

GzLMM:

ln_mmax_pm25		0.03787	0.01901	5915	1.99	0.0464

GEE MBSE:

ln_mmax_pm25	0.0532	0.0193	0.0155	0.0910	2.76	0.0057

GEE Empirical:

ln_mmax_pm25	0.0532	0.0151	0.0237	0.0828	3.53	0.0004

The estimate from the GzLMM is pretty different from the GEE (about 30% smaller), but the standard error from the GzLMM is very close to the SE from the model-based GEE approach. The SE from the empirical GEE approach is smaller than both of the other approaches.

e. Interpretation

The SD of the pollution variable is 0.592, so for each 1 SD increase in pollution, the rate of children's albuterol use increases by 2.27% (95% CI: 0.04% - 4.55%, p = 0.0464).

Questions

1.

c.

What are the drawbacks to Guassian quad/approximating the true likelihood?

Gaussian quad vs. ML?

d.

Drawbacks?

2.

b.

How do you compare apples to apples when fitting with RSPL? Or should I fit with MSPL?

e.

Do we need to go into detail about subject-specific vs. population-averaged interpretations?