STAA 554 Homework 1

Contents

Concept Questions

1. 5pts Nurse Stress Study.

Four wards were randomly selected at each of 25 hospitals and randomly assigned to offer a stress reduction program for nurses on the ward or to serve as a control. At the conclusion of the study period, a random sample of 10 nurses from each ward completed a test to measure job-related stress. Factors assumed to be related include nurse experience, age, hospital size and type of ward.

> **a.) Identify the level 1 observational units (most specific)**
>
> **b.) Identify the grouping units (could be multiple levels of grouping)**
>
> **c.) State the response(s) measured and variable type (normal, binary, Poisson, etc.)**
>
> **d.) Write a sentence describing the within-group correlation.**
>
> **e.) Identify all fixed and random effects in the model.**

a:The Individual nurse who is completing the test

b: The grouping units are the ward level and the hospital level

c: THe response is the stress level measured by a test. this is a continuous variable and should be normally distributed

d: nurses within the same ward tend to have more similar stress scores than nurses from different wards (and hospitals) as such there is within group correlation in stress levels

e: Fixed: Treatment, Nurse level excoriates, hospital level covariates

e: Random: random intercepts for ward random intercept for hospital

2. 5pts Radon.

Radon is a carcinogen - a naturally occurring radioactive gas whose decay products are also radioactive – known to cause lung cancer in high concentrations. The EPA sampled more than 80,000 homes across the US. Each house came from a randomly selected county and measurements were made on each level of each home. Uranium measurements at the county level were included to improve the radon estimates

> **a.) Identify the level 1 observational units (most specific)**
>
> **b.) Identify the grouping units (could be multiple levels of grouping)**
>
> **c.) State the response(s) measured and variable type (normal, binary, Poisson, etc.)**
>
> **d.) Write a sentence describing the within-group correlation.**
>
> **e.) Identify all fixed and random effects in the model.**

a: Measurement of radon taken on a floor of a home.

b: the house level and the county level are both grouping units

c: The response is the radon measurement which is continuous and non negative. It is also probably skewed so we may look at this as a normal or logNormal variable.

d: Measurements taken from different levels within the same home are likely to be more similar to each other than to measurements from different homes, because of building characteristics.

e: Fixed: County level uranium measurement, Floor level covariates like what floor of the house

e: Random: random intercept for home, random intercept for counties

Analysis: School Achievement Data

A researcher was interested in determining the extent to which vocabulary scores (`gevocab`) can be used to predict general reading achievement (`geread`). Students were nested within schools, (`school`) so that standard linear regression models would not be appropriate. In this case, school is a random effect, while vocabulary scores are fixed.

3. EDA

> **A) 1 pts. Table: student level**
>
> Use the `table1` command to create a summary table of the geread, gevocab, age and gender information (all measured at the student level).
> *Gender is coded as 1: Female, 2: Male*
> *Age is in months*

	Overall
	(N=10320)
geread	
Mean (SD)	4.34 (2.33)
Median [Min, Max]	3.80 [0, 12.0]

	Overall
gevocab	
Mean (SD)	4.49 (2.37)
Median [Min, Max]	3.80 [0, 11.2]
age	
Mean (SD)	108 (5.06)
Median [Min, Max]	107 [82.0, 135]
gender	
Female	5143 (49.8%)
Male	5177 (50.2%)

B) 1 pts. Table: school level

Use the `table1` command to create a summary table of ses (socio-economic level), and senroll (school enrollment) information (all measured at the school level). Hint: use the `gsummary()` function from the `nlme` package.

	Overall
	(N=160)
ses	
Mean (SD)	71.2 (23.0)
Median [Min, Max]	80.7 [0, 100]
senroll	
Mean (SD)	477 (162)
Median [Min, Max]	483 [115, 916]

4. Analysis

A) 2 pts.

Using `lmer()` fit the null model in which there is not an independent variable, but only the intercept and random effect. This model is useful for obtaining estimates of the residual and intercept variance when only the clustering by school is considered. Print a summary of your model below. We can obtain output from this model by typing `summary(Model)`.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: geread ~ 1 + (1 | school)
##    Data: data
##
## REML criterion at convergence: 46268.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3229 -0.6378 -0.2138  0.2850  3.8812
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
```

```
##   school   (Intercept) 0.3915    0.6257
##   Residual             5.0450    2.2461
## Number of obs: 10320, groups:  school, 160
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  4.30675    0.05498   78.34
```

> **B) 2 pts.**
>
> Although this is a null model in which there is not an independent variable, it does provide some useful information that will help us understand the structure of the data. In particular, the null model provides estimates of the variance among the individuals σ^2, and among the clusters τ^2. In turn, these values can be used to estimate the ICC. Calculate the ICC here.

```
## [1] 0.07201577
```

> **C) 2 pts.**
>
> Interpret your ICC value calculated above. 7.2% of the total variance is a result of the school level differences. the rest is due to individual level variability

> **D) 2 pts.**
>
> Fit a model with vocabulary test score as the independent variable and school as a random effect using `lmer()`. Provide a summary of the model below.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: geread ~ gevocab + (1 | school)
##    Data: data
##
## REML criterion at convergence: 43137.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0823 -0.5735 -0.2103  0.3207  4.4334
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  school   (Intercept) 0.09978  0.3159
##  Residual             3.76647  1.9407
## Number of obs: 10320, groups:  school, 160
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 2.023356   0.049309   41.03
## gevocab     0.512898   0.008373   61.26
##
## Correlation of Fixed Effects:
##         (Intr)
## gevocab -0.758
```
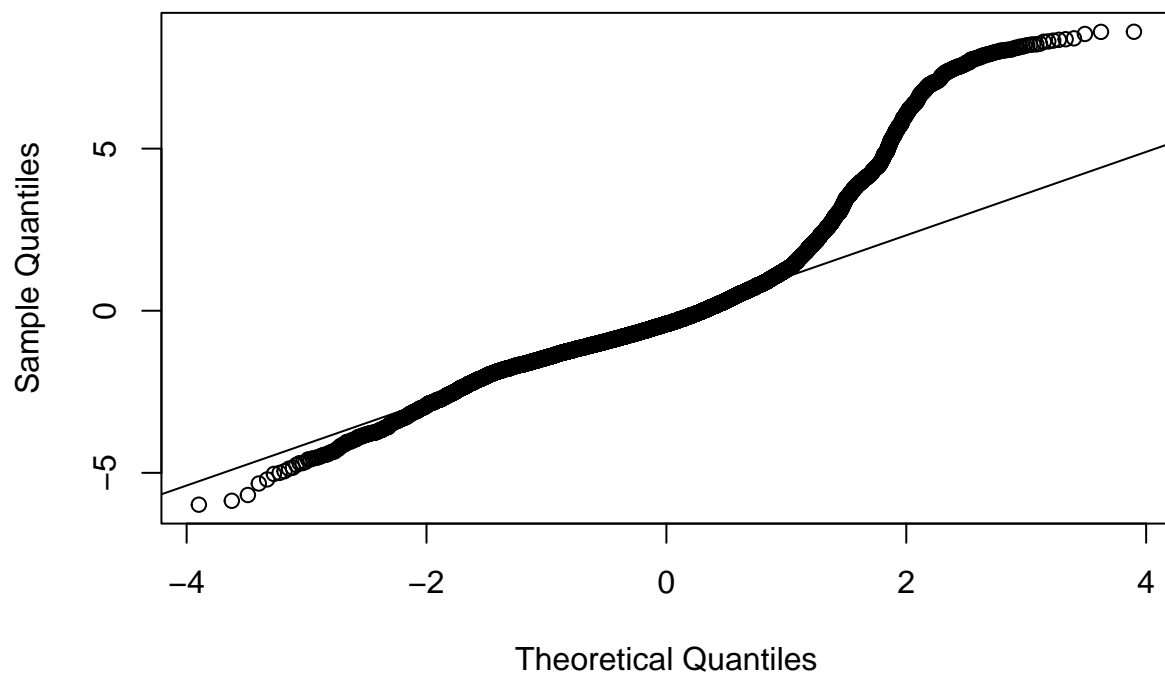
E) 2 pts.

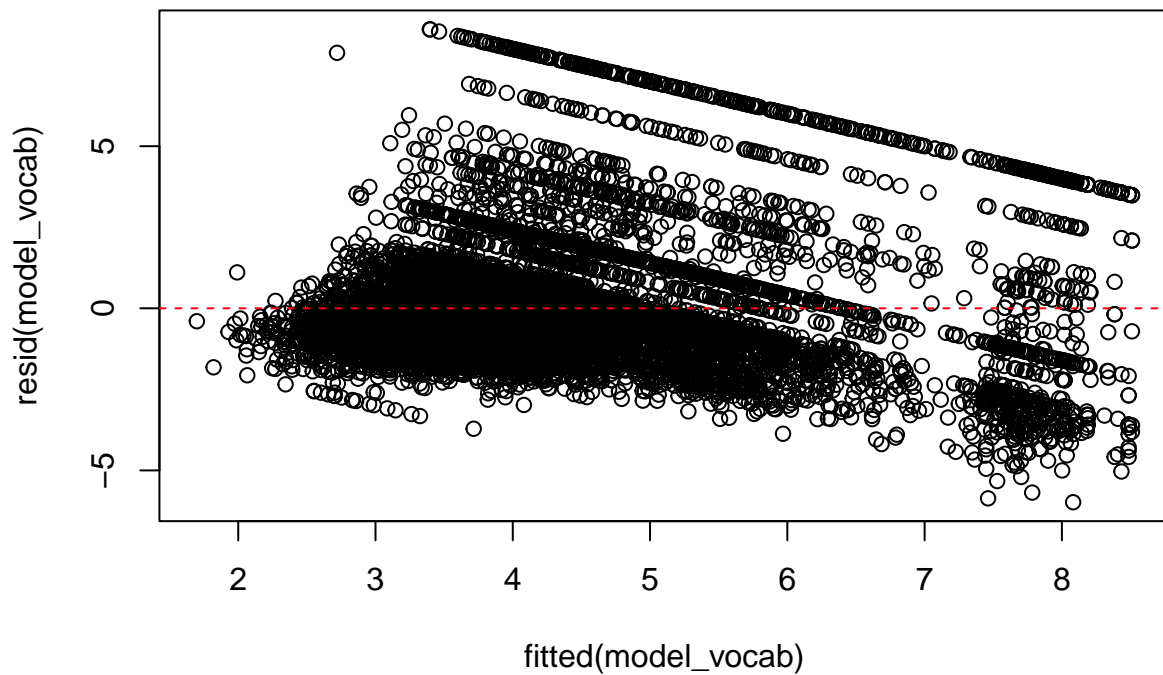Interpret the coefficient associated with `gevocab`.
For each point in vocabulary score, the reading score increases by about .51 points after accounting for inter school differences.

F) 4pts.

Assess the assumptions of normality and linearity with appropriate figures below. For each figure state your assessment of the assumption briefly.

Normal Q–Q Plot

The data does not pass normality which we see with the first residuals qq plot, and there is a clear issue if inequal variance as well assume clear stratification as seen in the fitted vs residual plot.

> **G) 4 pts.**
>
> Calculate and interpret the R^2 values for the level 1 model and the level 2 model (use the Snijders and Bosker definition).

```
## [1] 0.2534263
```

```
## [1] 0.7451475
```

References

Glasnapp, Douglas R., and John P. Poggio. Essentials of statistical analysis for the behavioral sciences. CE Merrill Pub. Co., 1985.
These notes are not to be posted or distributed anywhere without explicit consent of Dr. Eilertson.