

36-617: Applied Linear Models  
Fall 2021  
HW10 – Due Wed Nov 17, 11:59pm

- Please turn the homework in to Gradescope using the appropriate link in our course webspace at [canvas.cmu.edu](https://canvas.cmu.edu), under Assignments.
- No new reading.
- No quiz on Mon Nov 22.
- Please get started on this assignment right away. Problem #2 may take a while, for example (it's the technical appendix for Project 02!).

## Exercises

1. *Continuing with the Project 01 data...* The file `cdi.dat` in the Canvas folder for this hw is taken from Kutner et al. (2005): It provides selected county demographic information (CDI) for 440 of the most populous counties in the United States. Each line of the data set has an identification number with a county name and state abbreviation and provides information on 14 variables for a single county. Counties with missing data were deleted from the data set. The information generally pertains to the years 1990 and 1992. The definitions of the variables are given in Table 1 on p. 3 below. For this exercise we only consider a couple of these variables.

Use the function `rescale()` from `library(arm)` to rescale `pct.hs.grad` and `per.capita.income` as follows:

```
cdi$pci <- rescale(2*cdi$per.cap.income)
cdi$phg <- rescale(2*cdi$pct.hs.grad)
```

(This rescales the variables to have mean 0 and standard deviation 1; see `help(rescale)`. The rescaling is necessary because `lmer()` is a somewhat fragile function, and it can fail to converge correctly if the variables are on vastly different scales.)

- (a) Make a facet plot<sup>1</sup> like that on slide 7, lecture 22, using `state` as the grouping variable, with the following elements:
  - Scatter plot of  $x=phg$  and  $y=pci$  within each facet (state)
  - The fitted regression line for the completely pooled regression  $pci_i = \beta_0 + \beta_1 phg_i + \epsilon_i$ , ignoring states, plotted in each facet (state)
  - The fitted regression lines for the completely unpooled regression  $pci_i = \beta_{0[j[i]]} + \beta_{1[j[i]]} phg_i + \epsilon_i$  where  $j[i]$  is the state that observation  $i$  is in, and where  $\beta_{0j}$  and  $\beta_{1j}$  are fixed effects, plotted in each facet.
  - The fitted regression lines for the multilevel model

$$\begin{aligned} pci_i &= \alpha_{0[j[i]]} + \alpha_{1[j[i]]} phg_i + \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2) \\ \alpha_{0j} &= \beta_0 + \eta_{0j}, \quad \eta_{0j} \stackrel{iid}{\sim} N(0, \tau_0^2) \\ \alpha_{1j} &= \beta_1 + \eta_{1j}, \quad \eta_{1j} \stackrel{iid}{\sim} N(0, \tau_1^2) \end{aligned}$$

---

<sup>1</sup>Hint: see sample code in the file 22 - `mlm_residuals.r`.

and where the correlation between  $\eta_{0j}$  and  $\eta_{1j}$  is zero<sup>2</sup>.

- (b) Make a new facet plot with state as the grouping variable again, with x equal to the conditional  $\hat{y}$ 's, and y equal to the conditional residuals<sup>3</sup>, within each facet (state).

Submit your code, and the two plots. You do not need to write any commentary.

2. Make and submit sections of the technical appendix for your project 02 paper, using the hints and suggestions under **Suggestions for Answering the Research Questions** in the `project-02.pdf` assignment file:
  - (a) Section of your technical appendix for Research Question #1.
  - (b) Section of your technical appendix for Research Question #2.
  - (c) Section of your technical appendix for Research Question #3.
  - (d) Section of your technical appendix for Research Question #4.

---

<sup>2</sup>Hint: to force the correlation to be zero, rather than specifying the random effects as `(1 + phg | state)`, you should specify `(1 | state) + (0 + phg | state)`.

<sup>3</sup>Hint: `source("residual-functions.r")`, and use the appropriate functions defined in that file.

Table 1: Variable definitions for CDI data from Kutner et al. (2005). *Original source:* Geospatial and Statistical Data Center, University of Virginia.

| Variable Number | Variable Name                     | Description   |
|-----------------|-----------------------------------|---|
| 1               | Identification number             | 1–440   |
| 2               | County                            | County name   |
| 3               | State                             | Two-letter state abbreviation   |
| 4               | Land area                         | Land area (square miles)  |
| 5               | Total population                  | Estimated 1990 population   |
| 6               | Percent of population aged 18–34  | Percent of 1990 CDI population aged 18–34   |
| 7               | Percent of population 65 or older | Percent of 1990 CDI population aged 65 or older   |
| 8               | Number of active physicians       | Number of professionally active nonfederal physicians during 1990   |
| 9               | Number of hospital beds           | Total number of beds, cribs, and bassinets during 1990  |
| 10              | Total serious crimes              | Total number of serious crimes in 1990, including murder, rape, robbery, aggravated assault, burglary, larceny-theft, and motor vehicle theft, as reported by law enforcement agencies                      |
| 11              | Percent high school graduates     | Percent of adult population (persons 25 years old or older) who completed 12 or more years of school  |
| 12              | Percent bachelor's degrees        | Percent of adult population (persons 25 years old or older) with bachelor's degree  |
| 13              | Percent below poverty level       | Percent of 1990 CDI population with income below poverty level  |
| 14              | Percent unemployment              | Percent of 1990 CDI population that is unemployed   |
| 15              | Per capita income                 | Per-capita income (i.e. average income per person) of 1990 CDI population (in dollars)  |
| 16              | Total personal income             | Total personal income of 1990 CDI population (in millions of dollars)   |
| 17              | Geographic region                 | Geographic region classification used by the US Bureau of the Census, NE (northeast region of the US), NC (north-central region of the US), S (southern region of the US), and W (Western region of the US) |