Seeking Professional Help on Child Mental Health

A comparative study between a hierarchical Bayesian model and Classical Regressions

Samuel Annan, Huan Qin, and Piaomu Liu



Outline

- Background information and inspiration of study
- Data information
- Bayesian model
- Model estimation
- Frequentist model
- Model estimation
- Model comparison
- Discussion
- Conclusions

Background Information& Study Inspiration

- Inspired by previous studies on disparities of health outcomes: under-consumption of professional mental health services in the U.S. among minorities
- Our research question: by modeling national survey data, would we reach the same conclusion? How would other ethnic groups behave with respect to consumption of professional mental health service? How does the Bayesian methodology compare with the classical regression method; and how do both compare with reality (based on available data)?
- Measure of consumption of professional mental health service: the "visit" variable

Data

- National Health Interview Survey data downloaded from ICPSR-The Interuniversity Consortium for Political and Social Research website
- Participants were interviewed for behavioral statistics pertaining to general health. In particular we are interested in children's mental health.
- The aim is to analyze based on race to check whether the likelihood of visiting mental health professionals varies according to race in a way that provides information. Research suggests the existence of such variability by race.
- We used NHIS 2001 which gives the analysis information on children's mental health situations

Data (continued)

- The Sample Child file (from NHIS2001): 8,958 households were interviewed in the survey (we kept 8,598 observations for our analysis)
- SAS data step/procedures were used to prepare data for analysis
- Variables of interest include mother's education ("medu"7 levels), Parent presence at home("presence"4 levels), historical visit to mental health professionals ("visit"2 levels) and race ("hiscod_i"4 levels).
- The demographic summary of the data shows a 26.52% Hispanic, 53.15% White, 16.25% Black and 4.07% Others.
- More detailed descriptions on National Health Interview Survey can be found on the ICPSR website

• Data Source and Disclaimer

- Our analysis tries to capture as much information as can be collected to answer questions of interest to the extent to which the data is correct and void of bias.
- While migrating data via SAS procedures and modifying the data for use in Winbugs and MS Excel, steps were taken to preserve the integrity of the data.
- Data source: National Health Interview survey (NHIS), 2001 *http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/3605?q=national+health+interview*




Bayesian Model

$$V_i|p_i \sim bernoulli(p_i)$$

$$logit(p_i) = \beta_{oi} + \beta_{1i} * medu_i + \beta_{2i} * poverty_i + \beta_{3i} * presence_i + \varepsilon_i$$

- $\begin{pmatrix}\beta_{oi}\\ \beta_{1i}\\ \beta_{2i}\\ \beta_{3i}\end{pmatrix} \sim BVN\left(\mu = \begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\\ \mu_4\end{pmatrix}, \Sigma = \begin{pmatrix}\tau_{11} & \tau_{12} & \tau_{13} & \tau_{14}\\ \tau_{21} & \tau_{22} & \tau_{23} & \tau_{24}\\ \tau_{31} & \tau_{32} & \tau_{33} & \tau_{34}\\ \tau_{41} & \tau_{42} & \tau_{43} & \tau_{44}\end{pmatrix}\right)$
- $\varepsilon_i \sim N(0, \sigma^{\,2})$
- $\Sigma^{-1} \sim Wishart(R, k)$; R is a scaled matrix and k is the degree of freedom.
- $\mu_i \sim N(mean_i, prec_i)$
- $mean_i \sim N(0,1)$

Winbugs Codes

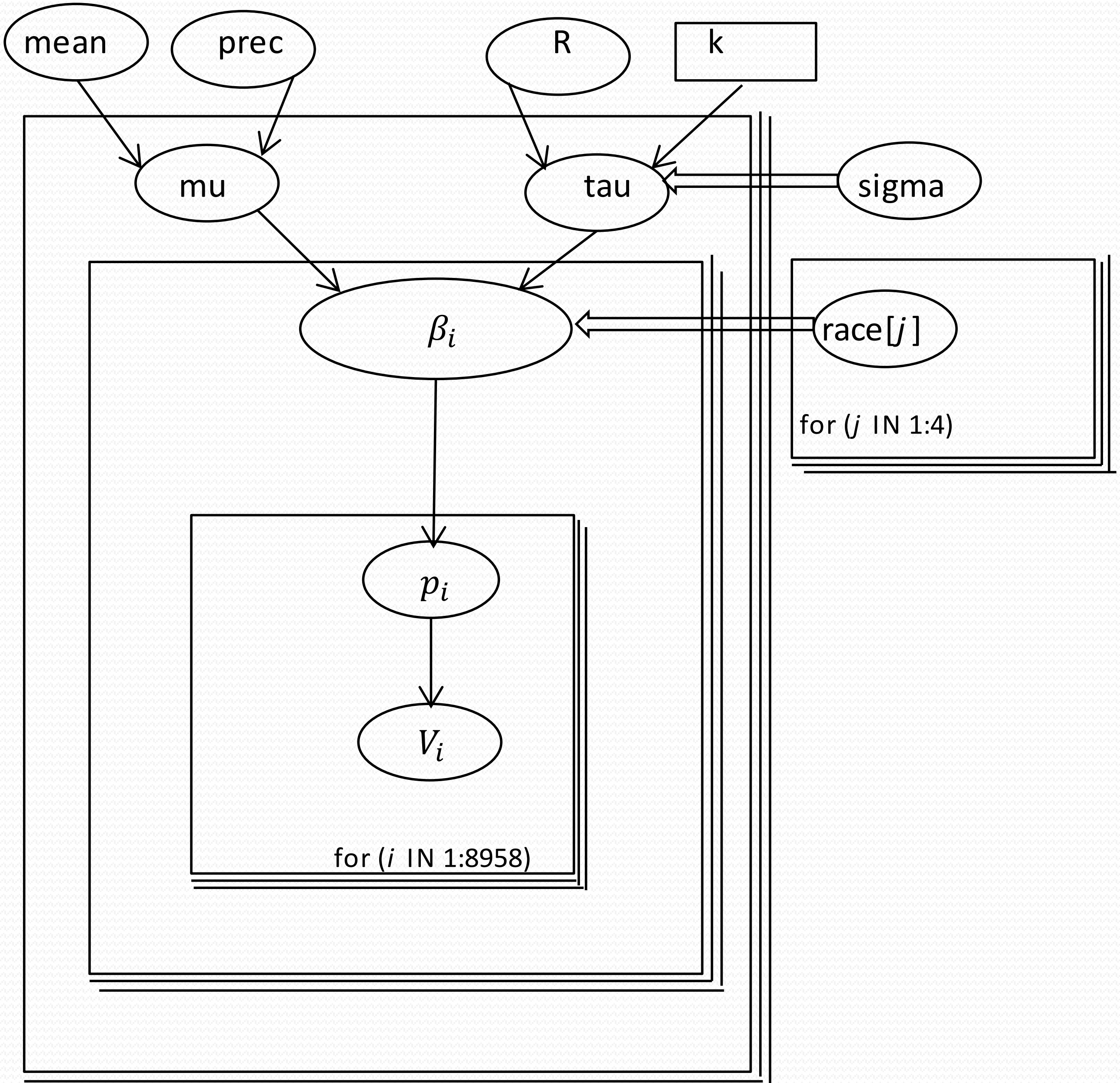
```
model
{
for (i in 1:8958){
        visit[i]~dbern(p[i]);
        logit(p[i])<-beta[race[i],1] + beta[race[i],2]*medu[i]  + beta[race[i],3]*poverty[i] + beta[race[i],4]*presence[i]
}
for (j in 1:4)
{
beta[j,1:4] ~ dmnorm(mubeta[],tau[,])
}
mubeta[1:4]~dmnorm(mean[],prec[,])
tau[1:4,1:4]~dwish(R[,],6)
sigma[1:4,1:4]<-inverse(tau[1:4,1:4])
mean[1] ~ dnorm(0, 1)
mean[2] ~ dnorm(0, 1)
mean[3] ~ dnorm(0, 1)
mean[4] ~ dnorm(0, 1)
}

list(prec=structure(.Data=c(1,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,1),.Dim=c(4,4)),
   R=structure(.Data=c(1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1),.Dim=c(4,4)))

race[]    visit[]    medu[]    poverty[]    presence[]
3          0          5          3            4
2          0          2          7            3
.          .          .          .            .
.          .          .          .            .
END

inits
list(mubeta=c(0,0,0,0), tau=structure(.Data=c(1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1),.Dim=c(4,4))
```

Directed graphs



Model Estimation

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1,1]	-2.945	0.2959	0.007872	-3.537	-2.939	-2.377	1001	20000
beta[1,2]	0.2672	0.08373	0.00182	0.09798	0.268	0.4287	1001	20000
beta[1,3]	0.08772	0.0629	0.001268	-0.03471	0.08768	0.2129	1001	20000
beta[1,4]	-0.4672	0.09702	0.00197	-0.6581	-0.4658	-0.2764	1001	20000
beta[2,1]	-1.756	0.2074	0.006529	-2.168	-1.752	-1.356	1001	20000
beta[2,2]	0.1153	0.04926	0.001361	0.01904	0.1155	0.2123	1001	20000
beta[2,3]	0.006562	0.03318	8.54E-4	-0.05789	0.006629	0.07143	1001	20000
beta[2,4]	-0.4127	0.06135	0.001539	-0.533	-0.4123	-0.2936	1001	20000
beta[3,1]	-3.791	0.4095	0.01243	-4.634	-3.781	-3.024	1001	20000
beta[3,2]	0.2785	0.1109	0.003083	0.05952	0.2791	0.4957	1001	20000
beta[3,3]	-0.09409	0.07213	0.00147	-0.2355	-0.09377	0.04739	1001	20000
beta[3,4]	0.1498	0.1174	0.002604	-0.07933	0.1496	0.386	1001	20000
beta[4,1]	-2.683	0.6643	0.01613	-4.034	-2.674	-1.379	1001	20000
beta[4,2]	0.5138	0.2015	0.004359	0.117	0.5115	0.9232	1001	20000
beta[4,3]	-0.09872	0.1447	0.003019	-0.3848	-0.09874	0.1802	1001	20000
beta[4,4]	-0.7409	0.2645	0.005588	-1.282	-0.735	-0.2347	1001	20000
mubeta[1]	-2.522	0.5118	0.007634	-3.446	-2.554	-1.384	1001	20000
mubeta[2]	0.2695	0.2679	0.002665	-0.2739	0.2733	0.8026	1001	20000
mubeta[3]	-0.02162	0.2534	0.002209	-0.5297	-0.02149	0.4786	1001	20000
mubeta[4]	-0.3929	0.3031	0.003248	-1.019	-0.387	0.1907	1001	20000

Hispanic: $logit(p_i) = -2.945 + 0.2672 * medu_i + 0.0878 * poverty_i - 0.4672 * presence_i$

Whites: $logit(p_i) = -1.756 + 0.1153 * medu_i + 0.0066 * poverty_i - 0.4127 * presence_i$

Blacks: $logit(p_i) = -3.791 + 0.2785 * medu_i - 0.0941 * poverty_i + 0.1498 * presence_i$

Others: $logit(p_i) = -2.683 + 0.5138 * medu_i - 0.0987 * poverty_i - 2.5220 * presence_i$

Results

- For all race categories, poverty level was the most significant variable.
- For Blacks, in addition to poverty level, parental presence was also influential in the determination of a child's likelihood to visit a mental health professional.
- The probabilities for each of the 8598 observations were predicted using the model and compared with both reality and the results from the regression procedure.

Race	n	Likelihood
White	406	8.53%
Hispanic	109	4.59%
Black	72	4.95%
Others	15	4.11%

Proportion by Percentile				Decile	Proportion by race			
H	W	B	O		H	W	B	O
74.47%	0.00%	1.23%	24.30%	0.1	27.99%	0.00%	0.76%	59.45%
60.71%	0.00%	36.72%	2.57%	0.2	22.90%	0.00%	22.60%	6.30%
32.81%	0.00%	65.18%	2.01%	0.3	12.37%	0.00%	40.11%	4.93%
30.69%	36.38%	28.91%	4.02%	0.4	11.57%	6.85%	17.79%	9.86%
1.90%	98.10%	0.00%	0.00%	0.5	0.72%	18.46%	0.00%	0.00%
13.30%	80.56%	5.25%	0.89%	0.6	5.01%	15.14%	3.23%	2.19%
11.94%	88.06%	0.00%	0.00%	0.7	4.50%	16.57%	0.00%	0.00%
12.17%	74.78%	10.60%	2.46%	0.8	4.59%	14.07%	6.52%	6.03%
25.45%	58.82%	13.06%	2.68%	0.9	9.60%	11.07%	8.04%	6.58%
2.00%	94.54%	1.56%	1.89%	1	0.76%	17.83%	0.96%	4.66%

The Classical Regression Approach (Fixed Effects)

- Logistic regression with fixed effects:

$$logit\,(p_i) = \beta_{0i} + \beta_{1i} * presence_i + \beta_{2i} * poverty_i + \beta_{3i} * poverty_i + \beta_{4i} * race_i * presence_i + \varepsilon_i$$

$$\eta_i = \beta_{0i} + \beta_{1i} * presence_i + \beta_{2i} * poverty_i + \beta_{3i} * poverty_i + \beta_{4i} * race_i * presence_i + \varepsilon_i$$

$$p_i = \exp(\eta_i) \big/ (1 + \exp(\eta_i))$$

- **SAS codes**:

```
proc logistic data=temp;
class hiscod_i presence poverty;
model visit = hiscod_i  poverty presence
hiscod_i*presence/ctable;
output out=p2 predicted = prob xbeta=logit;
run;
```

The Classical Regression Approach (random effects)

- Logistic regression with random effects :

$Visit_{ij} \sim Bernoulli(P_{ij})$

$logit\ (P_{ij}) = \beta_0 + \beta_{1j} * presence + \beta_{2j} * poverty + \mu_{ij}$

$\eta_{ij} = \beta_0 + \beta_{1j} * presence + \beta_{2j} * poverty + \mu_i$

$\mu_{ij} \sim Normal\ (0, \sigma)$

$$P_{ij} = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})}$$

- **SAS codes:**

```
proc nlmixed data=temp;
parms beta0=-1 beta1=1 beta2 = 1 s2u=2;
eta    = beta0 + beta1*presence + beta2*poverty + u;
expeta = exp(eta);
p = expeta/(1+expeta);
model visit ~ binary(p);
random u ~ normal(0,s2u) subject=hiscod_i;
predict eta out=eta;
run;
```

Regression Parameter Estimates

The fixed effects model

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
HISCOD_I	3	5.6954	0.1274
poverty	6	11.2367	0.0813
presence	4	28.3235	<.0001
HISCOD_I*presence	9	34.9928	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	8.5946	175.5	0.0024	0.9609
HISCOD_I	1	1	-2.1733	67.4128	0.0010	0.9743
HISCOD_I	2	1	-3.0574	67.4123	0.0021	0.9638
HISCOD_I	3	1	-3.4359	67.4123	0.0026	0.9594

The random effects model

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta0	-7.5271	0.2988	5579	-25.19	<.0001	0.05	-8.1129	-6.9413	0.000244
beta1	-0.5259	0.08686	5579	-6.05	<.0001	0.05	-0.6962	-0.3556	0.002669
beta2	0.04588	0.04567	5579	1.00	0.3151	0.05	-0.04364	0.1354	-0.00182
s2u	89.4317	10.1491	5579	8.81	<.0001	0.05	69.5356	109.33	7.345E-6

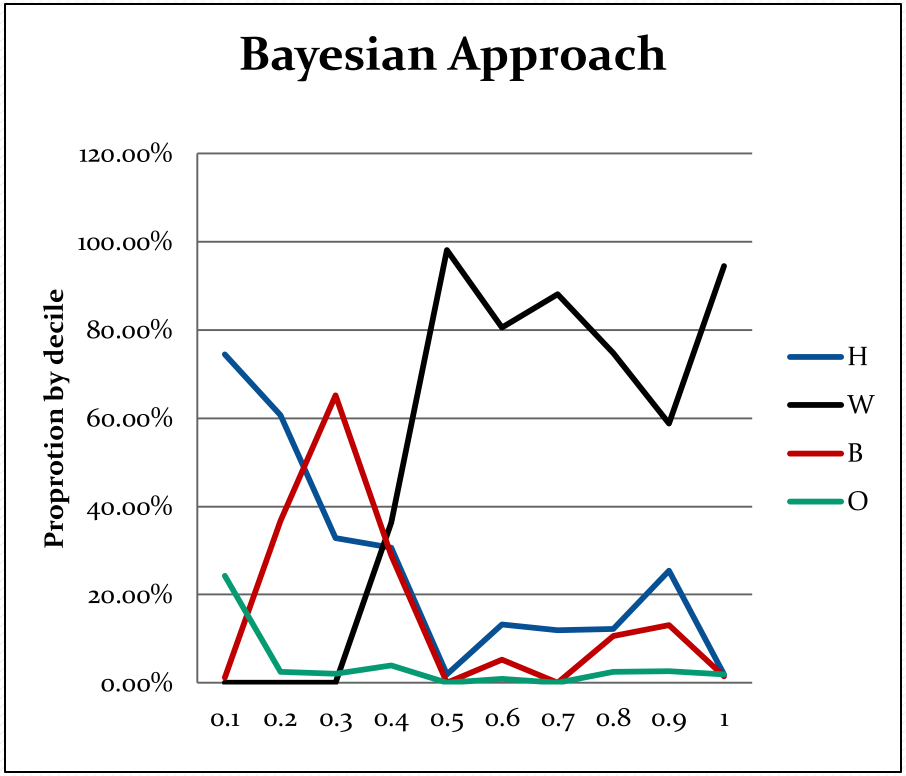
Results (the classical regression approach)

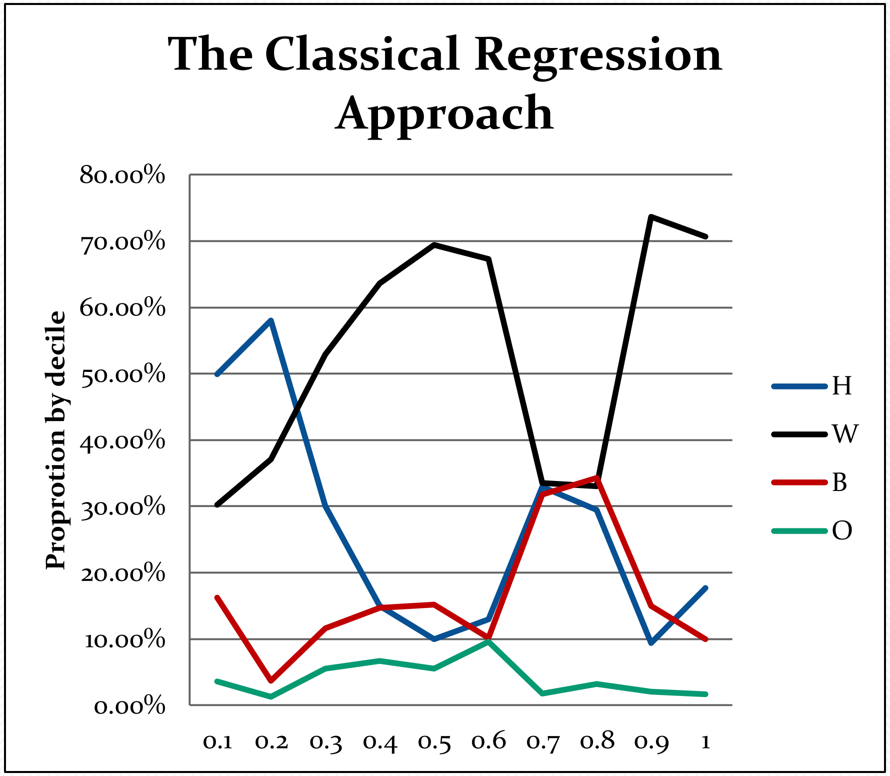
- The fixed effect model with interaction between the race variable and presence: controlling for poverty, both race and race * presence are significant
- The random effect model: both race and presence are significant, but not poverty
- The probabilities for each of the 8598 observations were predicted using the model and compared with both reality and the results from the regression procedure

Race	n	Likelihood
White	406	8.53%
Hispanic	109	4.59%
Black	72	4.95%
Others	15	4.11%

Proportion by Percentile				Deciles	Proportion by race			
H	W	B	O		H	W	B	O
49.94%	30.24%	16.24%	3.58%	0.1	18.77%	5.67%	9.96%	8.77%
58.04%	37.05%	3.68%	1.23%	0.2	21.89%	6.97%	2.27%	3.01%
30.02%	52.90%	11.61%	5.47%	0.3	11.32%	9.96%	7.14%	13.42%
14.96%	63.62%	14.73%	6.70%	0.4	5.64%	11.97%	9.07%	16.44%
9.93%	69.42%	15.18%	5.47%	0.5	3.75%	13.06%	9.34%	13.42%
12.96%	67.26%	10.17%	9.61%	0.6	4.88%	12.64%	6.25%	23.56%
32.92%	33.48%	31.81%	1.79%	0.7	12.42%	6.30%	19.57%	4.38%
29.46%	33.04%	34.26%	3.24%	0.8	11.11%	6.22%	21.09%	7.95%
9.38%	73.66%	14.96%	2.01%	0.9	3.54%	13.86%	9.20%	4.93%
17.71%	70.71%	9.91%	1.67%	1	6.69%	13.34%	6.11%	4.11%

- Comparisons





Discussion

- Results were similar
  - Both reported high likelihoods for whites followed by Hispanics and Blacks, and Others (in order).
  - Both results agree with frequencies gathered from data and are consistent with literature (to be discussed further)
  - However, conclusion about significance of the variable “poverty” are different

Conclusions & Discussions

- Likelihood of child seeking mental health help vary from race to race
- Both Bayesian and the classical regression methods yield results that are comparable
- Consistent with the data (Sample Child file) , our results are also consistent with previous studies in psychology
- The analysis may be improved by the following (including but not limited to):

  1. The Bayesian model: sensitivity analysis (it took Winbugs a long time to run……); change the link function of the Bayesian model
  2. The regression model: under certain conditions, different procedures in SAS can build same models; the analysis will be able to construct hierarchical model using PROC GILIMMIX if data have better representation of different racial groups
  3. Different groupings of categorical independent variables may lead to different results from analysis

Appendix & References

Appendix

- visit(our y variable): 0 (didn't see a mental health professional); 1 otherwise
- hiscod_i(our race variable): 1 Hispanic, 2 white, 3 black, 4 others
- poverty (income to poverty ratio): 1 lower than .5,……etc the higher the number is, the better the financial situation of the household; 15 "unknown"
- medu(mother's education): the higher the number is, the better
- presence (if parents are present in the household): 1 Mother not father 2 Father not mother 3 Mother and father 4 Neither 5 Professional degree 6 unknown

Reference

- Francesca Dominici, PhD www.biostat.jhsph.edu/~fdominic/
- Mary Kathryn Cowles, PhD

  www.stat.uiowa.edu/~kcowles

Thank you! & Questions???