References


What do we have so far?

- Point process ideas (intensities, $K$-functions).
  - Data: $(x, y)$ event locations.
  - Where are the clusters? Use intensities.
  - How are events clusters? Use $K$-functions.
- Disease clustering with point data.
- Disease clustering with regional counts.
What’s left?

- So we know how to describe and evaluate spatial patterns in health outcome data.
- What about linking patterns in health outcomes to patterns in exposures?
- With *independent* observations we know how to use *linear* and *generalized linear* models such as linear, Poisson, logistic regression.
- What happens with *dependent* observations?
"…all models are wrong. The practical question is how wrong do they have to be to not be useful."

*Box and Draper (1987, p. 74)*
What changes with dependence?

- In statistical modeling, we are often trying to describe the mean of the outcome as a function of covariates, assuming error terms are mutually independent.
- That means we usually model any trend in the data as a trend in *expectations*.
- Allows estimation of covariate effects.
- With *dependent* error terms, observed trends may be due to covariates, correlation, or both.
- May impact the identifiability of covariate effects.
- Could have different effects equally likely under different correlation models.
Residual correlation

- Where do correlated errors come from?
- Perhaps outcomes truly correlated (infectious disease).
- Perhaps we omitted an important variable that has spatial structure itself.
- If temperature is important and we left it out of a model applied to the continental U.S., what would the residuals look like?
If high temperatures associated with high outcomes, we would *underfit* in southern states (observations $\geq$ model predictions $\Rightarrow$ positive residuals), and *overfit* in northern states (observations $<$ model prediction $\Rightarrow$ negative residuals).

The “missing covariate” idea suggests that *maps* of residuals are important spatial diagnostics.

Also, we may want to apply tests of clustering or to detect clusters to residuals.

Moran’s $I$, LISAs.
NY leukemia data and add some covariates.

We will fit linear and Poisson regression models with various spatial correlation structures and compare inferences.

Remember, all of these models are wrong, but some may be useful.
New York leukemia data from Waller et al. (1994)
281 census tracts (1980 Census).
8 counties in central New York.
1,057,673 people at risk.
Crude Rates (per 100,000)

Central New York Census Tracts, 1980
Leukemia rates 1978-1982

- 0.000000 - 0.000021
- 0.000022 - 0.000324
- 0.000325 - 0.000588
- 0.000589 - 0.001096
- 0.001097 - 0.006993
Building the model

- Let \( Y_i \) = count for region \( i \).
- Let \( E_i \) = expected count for region \( i \).
- \( x_{i,TCE} \) = inverse distance to TCE site.
- \( x_{i,65} \) = percent over age 65 (census).
- \( x_{i,home} \) = percent who own own home (census).
- The model:

\[
Y_i = \beta_0 + x_{i,TCE} \beta_{TCE} + x_{i,65} \beta_{65} + x_{i,home} \beta_{home} + \epsilon_i .
\]
Assumptions for regression

- The error terms, $\epsilon_i \sim N(0, \sigma^2)$;
- The data have a constant variance, $\sigma^2$;
- The data are uncorrelated (OLS) or have a specified parametric covariance structure (GLS);
Y normally distributed?

Histogram

Normal Q–Q Plot

Incidence Proportions

Frequency

Theoretical Quantiles

Sample Quantiles

0.000 0.003 0.006

0.000 0.004

0 50 100 150

−3 −1 0 1 2 3

0.000
\[ Z_i = \log \left( \frac{1000(Y_i + 1)}{n_i} \right). \]
Outliers, where are the top 3?
Linear Regression (OLS)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_0$ (Intercept)</td>
<td>-0.5173</td>
<td>0.1586</td>
<td>0.0012</td>
</tr>
<tr>
<td>$\hat{\beta}_1$ (TCE)</td>
<td>0.0488</td>
<td>0.0351</td>
<td>0.1648</td>
</tr>
<tr>
<td>$\hat{\beta}_2$ (% Age &gt; 65)</td>
<td>3.9509</td>
<td>0.6055</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\hat{\beta}_3$ (% Own home)</td>
<td>-0.5600</td>
<td>0.1703</td>
<td>0.0011</td>
</tr>
<tr>
<td>$\hat{\sigma}^2$</td>
<td>0.4318</td>
<td>277 df</td>
<td></td>
</tr>
</tbody>
</table>

$R^2 = 0.1932$  
AIC = 567.5
Is OLS appropriate?

- $Z$s roughly Gaussian (symmetric).
- Do $Z$s have constant variance?
- No, since population sizes vary.
- $\text{Var}(Z_i) = \text{Var} \left( \log \left( \frac{1000(Y_i + 1)}{n_i} \right) \right)$
- Try *weighted least squares* with weights $1/n_i$. 
# Linear Regression (WLS)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_0$ (Intercept)</td>
<td>-0.7784</td>
<td>0.1412</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\hat{\beta}_1$ (TCE)</td>
<td>0.0763</td>
<td>0.0273</td>
<td>0.0056</td>
</tr>
<tr>
<td>$\hat{\beta}_2$ (% Age &gt; 65)</td>
<td>3.8566</td>
<td>0.5713</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\hat{\beta}_3$ (% Own home)</td>
<td>-0.3987</td>
<td>0.1531</td>
<td>0.0097</td>
</tr>
<tr>
<td>$\hat{\sigma}^2$</td>
<td>1121.94</td>
<td>277 df</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.1977</td>
<td></td>
<td>AIC=513.5</td>
</tr>
</tbody>
</table>
What changed?

- The three outliers are all in regions with small $n_i$.
- Weighting reduced their impact on estimates.
- Most profound effect is with respect to TCE.
WLS fitted values

WLS Fitted Values
-0.953 - -0.565
-0.565 - -0.48
-0.48 - -0.342
-0.342 - -0.156
-0.156 - 1.198
Residual plot
What are we looking for?

- Patterns in locations of residuals.
- Model underfit (predictions too low) near cities?
- Correlations in residuals?
- Let’s try semivariograms for the residuals.
- Let’s try local Moran’s $I$ for residuals.
Residual correlation?
Residual semivariogram not too impressive.

We can try *maximum likelihood* fit incorporating residual correlation via the semivariogram (which defines covariance matrix).
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_0$ (Intercept)</td>
<td>-0.7222</td>
<td>0.1972</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\hat{\beta}_1$ (TCE)</td>
<td>0.0826</td>
<td>0.0434</td>
<td>0.0576</td>
</tr>
<tr>
<td>$\hat{\beta}_2$ (% Age &gt; 65)</td>
<td>3.7093</td>
<td>0.6188</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\hat{\beta}_3$ (% Own home)</td>
<td>-0.3245</td>
<td>0.2044</td>
<td>0.1136</td>
</tr>
</tbody>
</table>

$\hat{c}_0 = 0.3740$  $\hat{c}_s = 0.0558$  $\hat{\alpha} = 6.93$

AIC = 565.6  df = 277
Weighting?

- We also need to include weights to account for heteroskedasticity.
- Again we use weights equal to $1/n_i$.
- What changes?
### Linear regression, Correlated, Weighted

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_0$ (Intercept)</td>
<td>-0.9161</td>
<td>0.1648</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\hat{\beta}_1$ (TCE)</td>
<td>0.0956</td>
<td>0.0322</td>
<td>0.0032</td>
</tr>
<tr>
<td>$\hat{\beta}_2$ (% Age &gt; 65)</td>
<td>3.5763</td>
<td>0.5920</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>$\hat{\beta}_3$ (% Own home)</td>
<td>-0.2285</td>
<td>0.1761</td>
<td>0.1956</td>
</tr>
</tbody>
</table>

$\hat{c}_0 = 997.65$  
$\hat{c}_s = 127.12$  
$\hat{a} = 6.86$

$\text{AIC} = 514.7$  
$\text{277 df}$
Fitted values (correlated, weighted)
Modelling counts directly

- Using linear regression required a fair amount of data transformation, just to meet modelling assumptions.
- Can we model the counts directly?
- In epidemiology, common to use logistic or Poisson regression.
- For rare disease, little difference between logistic and Poisson.
- Both are examples of generalized linear models (McCullagh and Nelder, 1989).
Building the model

Let \( Y_i = \text{count for region } i \).

Let \( E_i = \text{expected count for region } i \).

Let \((x_i, TCE, x_{i,65}, x_{i,\text{home}})\) be the associated covariate values.

Poisson regression:

\[ Y_i \sim \text{Poisson}(E_i \zeta_i) \]

where

\[ \log(\zeta_i) = \beta_0 + x_iTCE \beta_{TCE} + x_{i,65} \beta_{65} + x_{i,\text{home}} \beta_{\text{home}}. \]
What’s different?

- Poisson distribution for counts, rather than transforming proportions for normality.
- Link function: Natural log of mean of $Y_i$ is a linear function of covariates.
- So $\beta$s represent multiplicative increases in expected counts, $e^\beta$ a measure of relative risk associated with one unit increase in covariate.
- $E_i$ an offset, what we expect if the covariates have no impact.
- Age, race, sex adjustments in either $E_i$ (standardization) or covariates.
How do we add spatial correlation?

▶ Trickier than in regression, since mean and variance are related for Poisson observations.
▶ Two general approaches:
  ▶ *Marginal specification* defining correlation among means.
  ▶ *Conditional specification* defining correlation through the use of *random effects*. 

We often think of a model representing the *marginal mean*, $E(Y)$ as a function of fixed, unknown parameters.

That is, the parameters define the *population average* effect of the covariates (“On average, how does a given level of air pollution impact a person?”)

Another approach is to consider a model of the *conditional mean* for each subject.

In this setting we think of fixed effects of parameters and *random* effects specific to the subjects.
For us: *fixed effects* apply equally to all subjects, *random effects* apply to a particular subject.

Interpret fixed effects *conditional on* levels of the random effects.

“What is the effect of aspirin on a headache averaged over all individuals in the study?” (Marginal effect).

“What is the effect of aspirin on a headache in this individual?” (Conditional effect).

Random effects allow different parameter values for individuals, following some distribution.
Random intercepts

- A model with fixed and random effects is a *mixed* model.
- A very common formulation is to have fixed parameter values and a *random intercept*. This says everyone has the same response to the treatment, but that individuals have different starting points.
- In Poisson regression setting, if we add random effects we generate a *generalized linear mixed model* (GLMM).
Random effects and the conditional specification

- We add a *random effect* (intercept).
- Represents an impact of region $i$, not accounted for in $E_i$ or the covariates.
- We define this random effect to have a *spatial* distribution.
Let $Y_i$ denote the observed number of cases in region $i$.

Let $E_i$ denote the expected number of cases, ignoring covariate effects.

Assume $E_i$ known, perhaps age-standardized, or based on global (external or internal) rates.

First stage:

$$Y_i | \zeta_i \overset{ind}{\sim} \text{Poisson}(E_i \zeta_i)$$

$\zeta_i$ represent a relative risk associated with region $i$ not accounted for by the $E_i$. 
Building the model

- Note $Y_i/E_i = SMR_i$, the MLE of $\zeta_i$.
- Also note, $E[Y_i|\zeta_i] \neq E_i$, since $E_i$ does not include the impact of the random effect.
- Create a GLMM with log link by

$$\log(E[Y_i|\zeta_i]) = \log(E_i) + \log(\zeta_i)$$

- If we add covariates and rename $\log(\zeta_i) = \psi_i$, then

$$\log(\zeta_i) = x_i^T \beta + \psi_i$$
So our model is

\[ Y_i | \beta, \psi_i \overset{\text{ind}}{\sim} \text{Poisson}(E_i \exp(x_i' \beta + \psi_i)), \]

\[ \log(\zeta_i) = \beta_0 + x_{i,TCE} \beta_{TCE} + x_{i,65} \beta_{65} + x_{i,\text{home}} \beta_{\text{home}} + \psi_i. \]

The \( \psi_i \) represent the random intercepts.

Add overdispersion via \( \psi_i \overset{\text{ind}}{\sim} N(0, \nu_\psi). \)

Add spatial correlation via

\[ \psi \sim \text{MVN}(\mathbf{0}, \Sigma). \]
Priors and “shrinkage”

- Overdispersion model (i.i.d. $\psi_i$) results in each estimate being a compromise between the local SMR and the global average SMR.
- “Borrows information (strength)” from other observations to improve precision of local estimate.
- “ Shrinks” estimate toward global mean. (Note: “shrink” does not mean “reduce”, rather means “moves toward”).
Spatial model (correlated $\psi_i$) results in each estimate begin a compromise between the local SMR and the local average SMR.

Shrinks each $\psi_i$ toward the average of its neighbors.

Can also include both global and local shrinkage (Besag, York, and Mollié 1991).

How do we fit these models?
Bayesian inference regarding model parameters based on *posterior distribution*

\[ Pr[\beta, \psi | Y] \]

proportional to the product of the likelihood times the prior

\[ Pr[Y | \beta, \psi] Pr[\psi] Pr[\beta]. \]

Defers spatial correlation to the prior rather than the likelihood.
Spatial priors

- Could model joint distribution
  \[ \psi \sim MVN(0, \Sigma). \]

- Could also model conditional distribution
  \[ \psi_i | \psi_{j \neq i} \sim N \left( \frac{\sum_{j \neq i} c_{ij} \psi_j}{\sum_{j \neq i} c_{ij}}, \frac{1}{v_{CAR} \sum_{j \neq i} c_{ij}} \right), \ i = 1, \ldots, N. \]
  where \( c_{ij} \) are weights defining the neighbors of region \( i \).

- Adjacency weights: \( c_{ij} = 1 \) if \( j \) is a neighbor of \( i \).
CAR priors

- The conditional specification defines the *conditional autoregressive* (CAR) prior (Besag 1974, Besag et al. 1991).
- Under certain conditions on the $c_{ij}$, the CAR prior defines a valid multivariate joint Gaussian distribution.
- Variance covariance matrix a function of the *inverse* of the matrix of neighbor weights.
Given the values of the random effects ($\psi_i$s), observations ($Y_i$s) are independent.

Taking into account correlation in the $\psi_i$s, the $Y_i$s are correlated.

Conditionally independent $Y_i|\psi_i$ give likelihood function.

(Spatially correlated) distribution of the $\psi_i$s a prior distribution.
Posterior often difficult to calculate mathematically.
Iterative simulation approach to model fitting.
Given *full conditional* distributions, simulate a new value for each parameter, holding the other parameter values fixed.
The set of simulated values converges to a sample from the posterior distribution.
WinBUGS software.
www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml
Suppose we have a model with data $Y$ and three parameters $\theta_1$, $\theta_2$, and $\theta_3$.

“Gibbs sampler” simulates values from the full conditional distributions

$$f(\theta_1|\theta_2, \theta_3, Y),$$

$$f(\theta_2|\theta_1, \theta_3, Y),$$

$$f(\theta_3|\theta_1, \theta_2, Y).$$
Conceptual MCMC

- Start with values $\theta_1^{(1)}$, $\theta_2^{(1)}$, and $\theta_3^{(1)}$.

  sample $\theta_1^{(2)}$ from $f(\theta_1|\theta_2^{(1)}, \theta_3^{(1)}, Y)$,

  sample $\theta_2^{(2)}$ from $f(\theta_2|\theta_1^{(2)}, \theta_3^{(1)}, Y)$,

  sample $\theta_3^{(2)}$ from $f(\theta_3|\theta_1^{(2)}, \theta_2^{(2)}, Y)$.

- As we continue to update $\theta$, sampled values become indistinguishable from a sample from the joint posterior distribution $f(\theta_1, \theta_2, \theta_3|Y)$. 
Gelman et al. (2004). Theoretical and MCMC results.

\[
\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \text{MVN} \left( \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).
\]

Uniform priors on \( \theta_1, \theta_2 \), yield posterior

\[
\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \sim \text{MVN} \left( \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).
\]
Full conditionals

- Multivariate results give \textit{full conditionals}

\[
\begin{align*}
\theta_1 | \theta_2, \mathbf{Y} & \sim N(Y_1 + \rho(\theta_2 - Y_2), 1 - \rho^2), \\
\theta_2 | \theta_1, \mathbf{Y} & \sim N(Y_2 + \rho(\theta_1 - Y_1), 1 - \rho^2).
\end{align*}
\]

- Let’s try a Gibbs sampler and compare to the theoretical results.
MCMC example

First 10 iterations

500 iterations

Theta1

Theta2

Density

Theta1

Theta2

Density
Almost custom-made for MCMC.
Defined for $\psi_i$, given $\psi_j$ for $j \neq i$.
We define neighborhood weights $c_{ij}$.
Complete model specification

\[ Y_i | \beta, \psi_i \overset{\text{ind}}{\sim} \text{Poisson}(E_i \exp(x_i' \beta + \psi_i)) , \]

\[ \log(\zeta_i) = \beta_0 + x_i,TCE \beta_{TCE} + x_i,65 \beta_{65} + x_i,\text{home} \beta_{\text{home}} + \psi_i . \]

\[ \beta_k \sim \text{Uniform}. \]

\[ \psi_i | \psi_{j \neq i} \sim N \left( \frac{\sum_{j \neq i} c_{ij} \psi_j}{\sum_{j \neq i} c_{ij}}, \frac{1}{\nu_{\text{CAR}} \sum_{j \neq i} c_{ij}} \right), \quad i = 1, \ldots, N. \]

\[ \frac{1}{\nu_{\text{CAR}}} \sim \text{Gamma}(0.5, 0.0005). \]
MCMC trace plots

Intercept

Age Effect

Exposure Effect

Home Ownership Effect

Iteration

Iteration

Iteration

Iteration
Posterior densities
## MCMC posterior estimates

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Posterior Median</th>
<th>95% Credible Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>0.048</td>
<td>(-0.355, 0.408)</td>
</tr>
<tr>
<td>$\beta_{65}$</td>
<td>3.984</td>
<td>(2.736, 5.330)</td>
</tr>
<tr>
<td>$\beta_{TCE}$</td>
<td>0.152</td>
<td>(0.066, 0.226)</td>
</tr>
<tr>
<td>$\beta_{home}$</td>
<td>-0.367</td>
<td>(-0.758, 0.049)</td>
</tr>
</tbody>
</table>
But there's more!

- A nifty thing about MCMC estimates:

  We get posterior samples from any function of model parameters by taking that function of the sampled posterior parameter values.

- Gives us posterior inference for $SMR_i = Y_{i,\text{fit}}/E_i$.

- Also can get $Pr[SMR_i > 200|\mathbf{Y}]$ and map these exceedence probabilities.
Posterior median SMRs

Posterior median local SMR
CAR prior

45.64 - 80
80.01 - 90
90.01 - 110
110.01 - 120
120.01 - 453.15
Posterior exceedence probabilities

Probability SMR > 150

Probability SMR > 200

Legend:
- 0 - 0.2
- 0.2 - 0.4
- 0.4 - 0.6
- 0.6 - 0.8
- 0.8 - 1
Example 2

Cryptozoology example

- County-specific reports of encounters with *Sasquatch* (Bigfoot).
- “...which brings us to the appropriateness of the Bigfoot example.”
- Data downloaded from www.bfro.net
- Sightings from counties in Oregon and Washington (Pacific Northwest).
- Probability of report related to population density?
- (Hopefully) rare events in small areas.
- Perhaps spatial smoothing will stabilize local rate estimates.
- Fit models with no random effects, exchangeable random effects, CAR random effects, convolution random effects.
Sasquatch Data

Number of Reports

Reports per 2000 Population

Legend

Population/Sq. Mi.

<table>
<thead>
<tr>
<th>Reports</th>
<th>Reports/Person</th>
<th>Population/Sq. Mi.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00000 - 0.00003</td>
<td>0.7 - 12.9</td>
</tr>
<tr>
<td>1-5</td>
<td>0.00003 - 0.00008</td>
<td>13.0 - 32.1</td>
</tr>
<tr>
<td>6-10</td>
<td>0.00008 - 0.00016</td>
<td>32.2 - 69.9</td>
</tr>
<tr>
<td>11-15</td>
<td>0.00016 - 0.00026</td>
<td>70.0 - 180.1</td>
</tr>
<tr>
<td>16-20</td>
<td>0.00026 - 0.00046</td>
<td>180.2 - 414.0</td>
</tr>
<tr>
<td>21-25</td>
<td>0.00046 - 0.00076</td>
<td>414.1 - 793.3</td>
</tr>
<tr>
<td>25-51</td>
<td>0.00076 - 0.00517</td>
<td>793.4 - 1419.3</td>
</tr>
</tbody>
</table>

2000 Population per Square Mile

King
Skamania
Wasco
Reports vs. Population Density

![Graph showing the relationship between reports and population density.](image-url)
Observed vs. Expected

Observed versus Expected Number of Reports

Skamania

King
Predicted relative risks and credible sets

Filled circle = Skamania, Filled square = Wasco
Mapped relative risks

Legend

0.00 - 1.00
1.01 - 2.00
2.01 - 3.00
3.01 - 4.00
4.01 - 15.00
approximately 70.00
Skamania Sasquatch Ordinances


- Big Foot Ordinance 69-1: “THEREFORE BE IT RESOLVED that any premeditated, willful and wonton slaying of any such creature shall be deemed a felony punishable by a fine not to exceed Ten Thousand Dollars ($10,000.00) and/or imprisonment in the county jail for a period not to exceed Five (5) years. ADOPTED this 1st day of April, 1969.”

- Big Foot Ordinance 1984-2:
  - Repealed felony and jail sentence.
  - Established a Sasquatch Refuge (Skamania County).
  - Clarified penalty (gross misdemeanor vs. misdemeanor) and penalty (fine and jail time), disallowed insanity defense, and clarified distinction between coroner designation of victim as humanoid (murder) or anthropoid (this ordinance).
Conclusions

- What method to use depends on what you want data you have and what question you want to answer.
- All methods try to balance trend (fixed effects) with correlation (here, with random effects).
- All models wrong, some models useful.
- Trying more than one approach often sensible.
- Few methods (including Monte Carlo simulation) in current GIS packages.