Uncertainty-Quantified Ranking of Restaurant Chains in Canada by Food Safety Compliance

Combining Data From Multiple Sources for Ranking

Michael Guerzhoy and Nathan Taback
{guerzhoy, nathan}@utstat.toronto.edu

CBC Television, April 2014

Nationwide Ranking by Food Safety Compliance

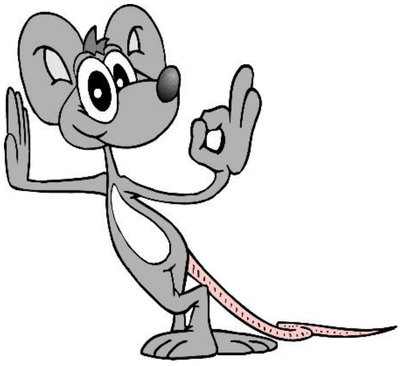
1. Starbucks
2. KFC
3. ...

UNIVERSITY OF TORONTO DCS50 CBCtelevision marketplace CANADA'S CONSUMER WATCHDOG WE'VE GOT YOUR BACK

Outline

- The problem: combining data from multiple cities to produce a nationwide ranking
- Selecting a ranking measure
- Quasi-Poisson regression for ranking
- Hierarchical Bayesian models for ranking
  - More differences between chains are identified
  - More flexible models
- Model building for exploring the dataset and producing a better ranking

The Problem



- Rank restaurant chains by food safety compliance
- Inspection reports on the number of major violations found during each inspection in 2013
  - Data from 5 Canadian cities
  - Inspectors use different standards in different cities
  - Average food safety levels may be different in different cities
- Need to quantify uncertainty

The Dataset (1)

- Data on food safety **inspections** of 100 of **stores** of 13 **chains** in 5 Canadian cities for 2013
- Approx. 3 inspections per store
- Number of major violations is recorded

1	Chain	Addr	Store	Report nc	Purpos	Date	Total Significant	TOTAL CRUCIAL	TEMPERATURE	FOOD HANDLING /HAND WASHING	KITCHEN CLEANLINESS	PEST CONTROL	CROSS-CONTAMINATION
2	Tim Hortons	2X2	T-T1	1	Routine	14-Mar-13	0						
3	Tim Hortons	2X2	T-T1	2	Routine	14-Mar-13	0						
4	Tim Hortons	2X2	T-T1	3	Routine	14-Dec-12	0						
5	Tim Hortons	6C7	T-T2	1	Routine	2-Oct-13	0						
6	Tim Hortons	6C7	T-T2	2	Routine	23-Jan-13	1				1		
7	Tim Hortons	6C7	T-T2	3	Routine	17-Sep-12	0						
8	Tim Hortons	0B6	T-T3	1	Routine	12-Jun-13	0						

The Dataset (2)

- (Very) different rates of violations in different cities
  - Rates of “major” violations differ by up to a factor of 4
  - Different standards in different cities?
- The numbers for Vancouver were assigned by an expert based on narrative inspection reports
- 2024 reports for Toronto, 1279 for Calgary, 877 for Ottawa, 472 for Vancouver, 118 for Regina

Ranking Chains in a Single City

- For each chain, compute the average/expected number of major violations found per inspection
- Standard errors are easily obtained

Halifax

Combining Data From Multiple Cities

- Data from Toronto, Vancouver, Calgary, Ottawa, and Regina
- The average number of violations per inspection differs by as much as a factor of four in different cities
  - Inspectors use different standards?
  - (Different average levels of compliance?)

Toronto *Ottawa* *Vancouver* *Calgary*

Ranking Measure

- Important to rank using a comprehensible measure!
- Rank by the *expected number of violations (using Toronto standards)* in a location of a given chain
- “If I visit a location of chain C, how many violations can I expect to encounter?”

Quasi-Poisson Regression:

$$N_{violations} \sim Poisson(\exp(c + a_{chain} + b_{city}))$$

- Unit of analysis: a single inspection visit
  - A random Canadian deciding where to go for lunch
- Model the expected number of violations as
  - $E(N_{violations}) = \exp(c + a_{chain} + b_{city})$ ⟵ log link fn
  - The larger a_{chain} , the more violations are assigned to ***chain***
  - The larger b_{city} , the more violations inspectors assign in ***city***
  - They combine multiplicatively, which makes sense

$$E(N_{violations}) = \exp(c + a_{chain} + b_{city})$$

- Rank by the number of violations that would be assigned in Toronto, based on all of the data:

  $$\exp(c + a_{chain} + b_{toronto})$$

  - The same as ranking by a_{chain}

  – "Expected number of violations using Toronto standards"

  – Standard errors for the *a* can be obtained by running quasi-Poisson regression on the data

  - Enables us to quantify uncertainty in the ranking
  - Standard errors adjusted for multiple comparisons using Tukey's Honest Significant Differences

	Major violations per inspection (adjusted to Toronto level) Non-adjusted figures given in square brackets	per 10	per 100	Worse than
Starbucks	0.11 (69.59/652) [153/652 non-adj.]	1	11	
KFC	0.13 (29.43/225) [44/225 non-adj.]	1	13	
A&W	0.17 (35.19/202) [86/202 non-adj.]	2	17	
Subway	0.18 (196.52/1087) [327/1087 non-adj.]	2	18	Starbucks
Pizza Hut	0.20 (29.20/147) [53/147 non-adj.]	2	20	
Tim Hortons	0.21 (213.06/994) [312/994 non-adj.]	2	21	Starbucks
Swiss Chalet	0.29 (66.48/229) [78/229 non-adj.]	3	29	KFC, Starbucks
Wendy's	0.30 (50.51/168) [83/168 non-adj.]	3	30	KFC, Starbucks, Subway
McDonald's	0.33 (160.46/487) [267/487 non-adj.]	3	33	A&W, KFC, Starbucks, Subway, Tim Hortons
Boston Pizza	0.36 (61.08/171) [117/171 non-adj.]	4	36	A&W, KFC, Starbucks, Subway, Tim Hortons
The Keg	0.37 (20.29/55) [32/55 non-adj.]	4		KFC, Starbucks
Second Cup	0.40 (105.70/263) [138/263 non-adj.]	4	40	A&W, KFC, Pizza Hut, Starbucks, Subway, Tim Hortons
Moxie's	0.49 (50.18/103) [104/103 non-adj.]	5	49	A&W, KFC, Pizza Hut, Starbucks, Subway, Tim Hortons

Table 4. Nationwide ranking: chains ranked by major violations per inspection, and the chains which are better than the given chain with 95% confidence. We report the estimated number of violations per inspection that would be assigned by Toronto inspectors to a location of a given chain. .

Report a set of significant differences at 95% confidence

A Bayesian Overdispersed Poisson Model

Check for overdispersion using fake-date simulation

$$city_i \sim N(0, \sigma_{city})$$

$$chain_k \sim N(0, \sigma_{chain})$$

$$\epsilon_j \sim N(0, \sigma_\epsilon)$$

overdispersion

$$\theta_j = \mu + city_{cities(j)} + chain_{chains(j)} + \epsilon_j$$

$$N\_VIOL_j \sim Poisson(exp(\theta_j))$$

(All priors are flat unless otherwise specified)

Ranking Using Samples from the Posterior Distribution

- Obtain samples from the joint posterior distribution of all the parameters using MCMC
- Rank chains by the median of the expected number of violations using Toronto standards

$$median(\sum_{c\in\{Toronto,Calgary,\ldots\}} P(c)\exp(\mu + chain_j + city_{Toronto}))$$

  - For *this* model, basically the same as ranking by the *chain* coefficients
  - P(c) is proportional to the population of city *c*

- Report differences whose joint probability is 95%

Uncertainty in the Rankings

- For a set of differences between chains, we can compute the joint probability that all the differences in the set hold
- Algorithm
  - Start with the list of all pairwise differences for (i, j) s.t. median(c_i)>median(c_j)
  - Remove differences from the list by order of significance (i.e., by P(c_i>c_j)) until the joint probability that all the differences in the list hold is >95%

	Per Inspection	Worse Than
Starbucks	0.07	
KFC	0.09	
A&W	0.11	Starbucks
Subway	0.12	Starbucks
Pizza Hut	0.13	Starbucks
Tim Hortons	0.14	KFC, Starbucks
Wendy's	0.19	A&W, Subway, KFC, Starbucks
Swiss Chalet	0.19	Tim Hortons, A&W, Subway, Starbucks, KFC
McDonald's	0.21	Pizza Hut, Starbucks, KFC, A&W, Subway, Tim Hortons
The Keg	0.22	A&W, Subway, Starbucks, KFC
Boston Pizza	0.22	Pizza Hut, Starbucks, KFC, A&W, Subway, Tim Hortons
Second Cup	0.25	Pizza Hut, Starbucks, KFC, A&W, Subway, Tim Hortons
Moxie's	0.26	Pizza Hut, Tim Hortons, Starbucks, KFC, A&W, Subway

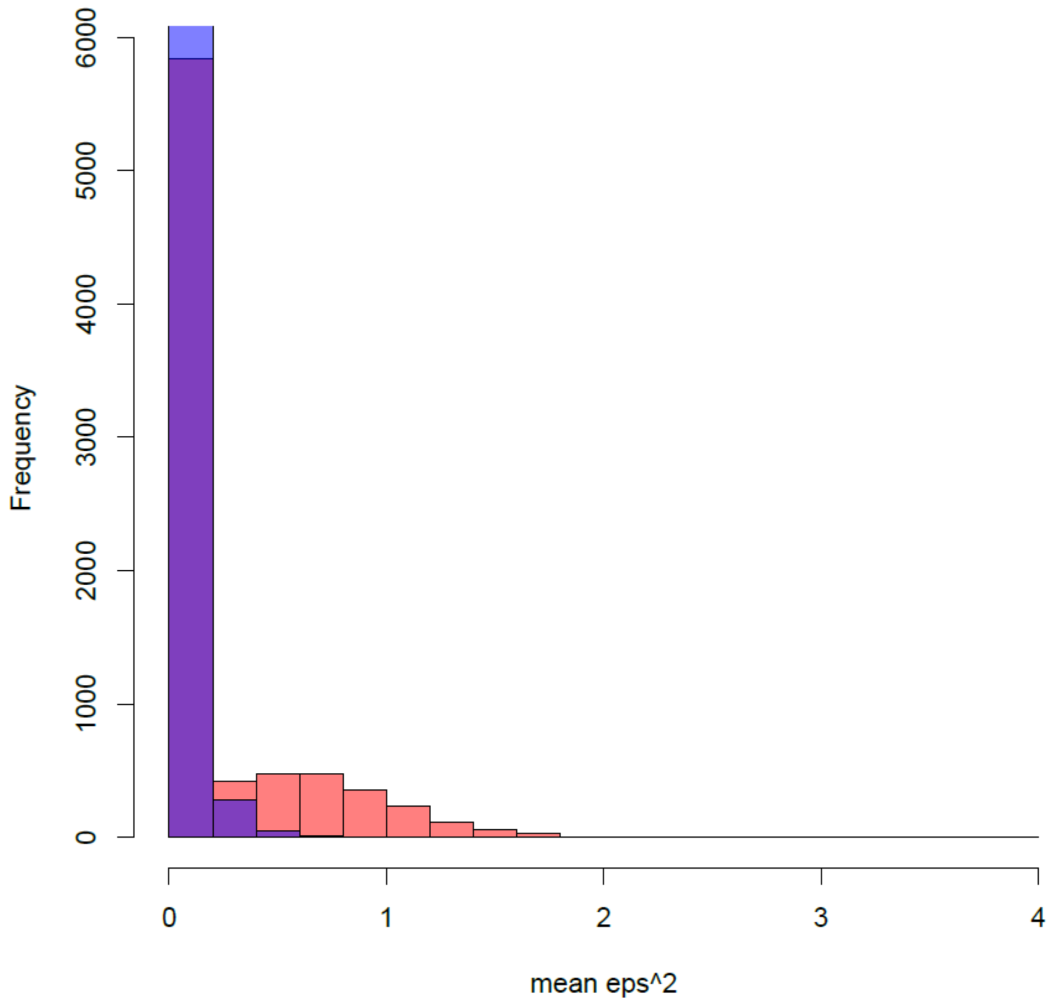
Bayesian vs. Frequentist methods

- More (41 vs. 31) differences are identified
  - The methods are not *quite* directly comparable, since the models are different
  - The rankings are almost the same

“Falsifying” the model

- Observation: data from Vancouver is substantially affecting the rank of Moxie’s
- Hypothesis: some chains are of uniform quality in all cities, and some aren’t

- Squared mean epsilon in locations belonging to the same [city, chain] pairs for actual posterior (red) and simulated (blue) parameters
- Histogram indicates systematic variation not accounted for by the city and chain parameters



$$city_i \sim N(0, \sigma_{city})$$

$$chain_k \sim N(0, \sigma_{chain})$$

$$\epsilon_j \sim N(0, \sigma_{\epsilon})$$

$$\theta_j = \mu + city_{cities(j)} + chain_{chains(j)} + \epsilon_j$$

$$N\_VIOL_j \sim Poisson(exp(\theta_j))$$

Variable Amount of Central Control

$$city_i \sim N(0, \sigma_{city})$$

$$chain_k \sim N(0, \sigma_{chain})$$

$$\epsilon_j \sim N(0, \sigma_{\epsilon})$$

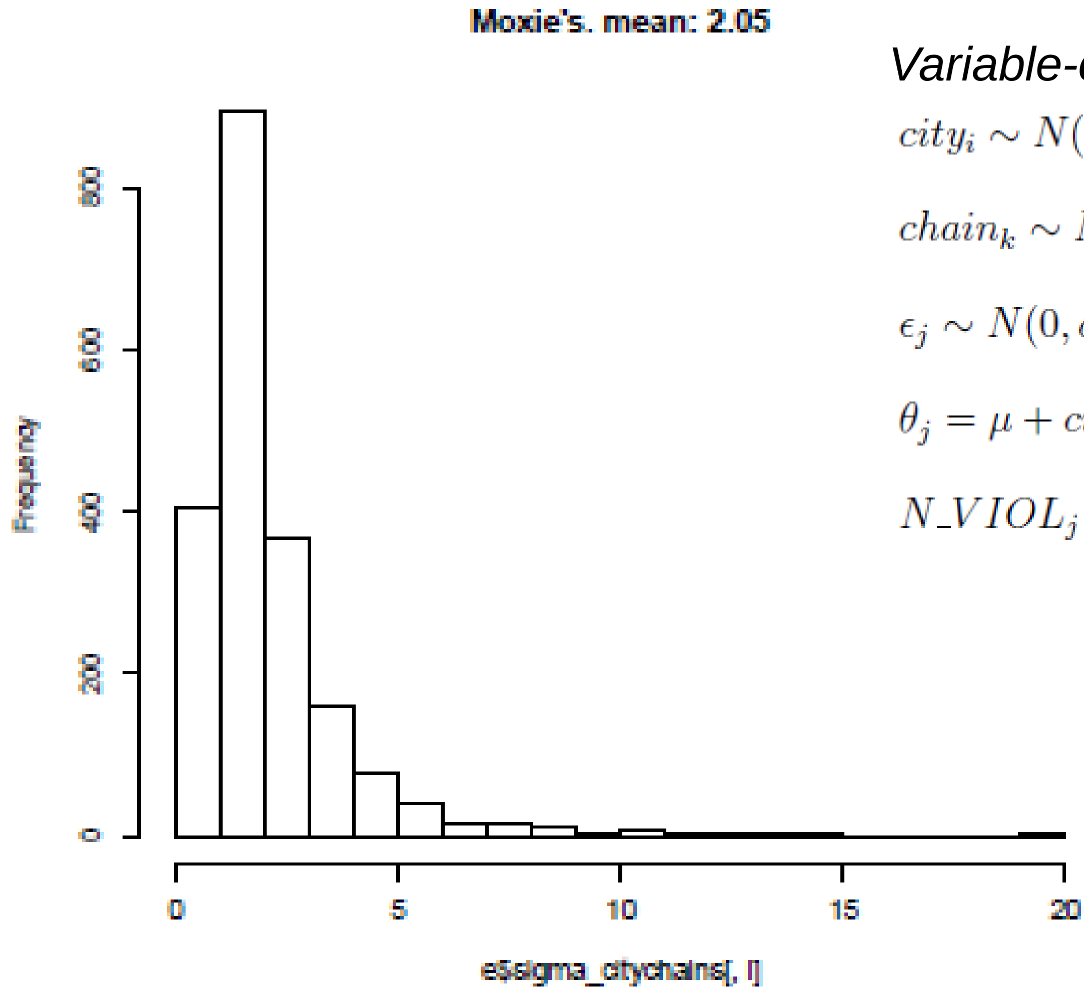
$$CITYCHAIN(i,k) \sim N(\mu + city_i + chain_k, \sigma_{citychain}(k))$$

$$\theta_j = CITYCHAIN(cities(j), chain(j)) + \epsilon_j$$

$$N\_VIOL_j \sim Poisson(exp(\theta_j))$$

If chain *k* is centrally controlled, $\sigma_{citychain}(k)$ is small

Does the Amount of Central Control Vary?



Variable-control model:

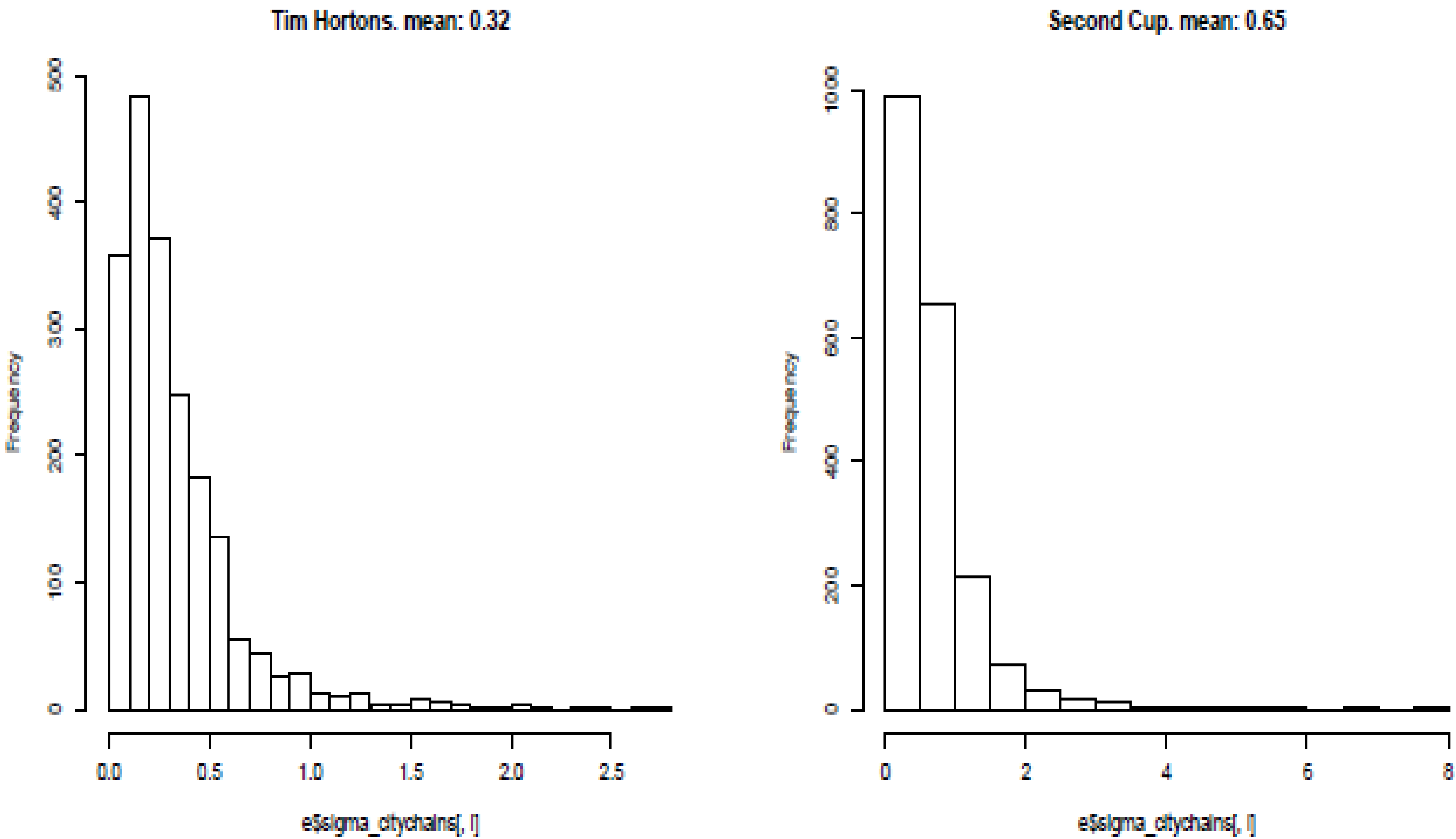
$$city_i \sim N(0, \sigma_{city})$$

$$chain_k \sim N(0, \sigma_{chain})$$

$$\epsilon_j \sim N(0, \sigma_\epsilon)$$

$$\theta_j = \mu + city_{cities(j)} + chain_{chains(j)} + \epsilon_j$$

$$N\_VIOL_j \sim Poisson(exp(\theta_j))$$



Tim Hortons. mean: 0.32
Frequency
0
100
200
300
400
500
0.0
0.5
1.0
1.5
2.0
2.5
e$sigma_citychains[, i]
Second Cup. mean: 0.65
Frequency
0
200
400
600
800
1000
0
2
4
6
8
e$sigma_citychains[, i]

Ranking using the Hierarchical Model

- Ranking using the *chain* coefficients alone would make sense if there had been a lot more major cities that weren’t sampled
  - In that case, sampling 5 cities would not be enough. Because of “explaining away” effects, there is a lot of uncertainty in the parameter estimates
- Ranking using the expected number of violations for a random diner in the 5 cities is still possible

	per Inspection	Worse than
Starbucks	0.07	
A&W	0.07	
KFC	0.10	
Subway	0.13	Starbucks
Tim Hortons	0.13	Starbucks
Pizza Hut	0.14	Starbucks
Wendy's	0.17	A&W, Starbucks
Swiss Chalet	0.20	Subway, KFC, Starbucks, A&W
McDonald's	0.22	Starbucks, A&W, KFC, Subway, Tim Hortons
Second Cup	0.23	Starbucks, A&W, KFC, Subway, Tim Hortons
The Keg	0.24	Subway, KFC, Starbucks, A&W
Moxie's	0.28	Tim Hortons, Starbucks, A&W, KFC, Subway
Boston Pizza	0.40	McDonald's, Wendy's, Starbucks, A&W, KFC, Subway, Tim Hortons, Pizza Hut

Conclusions

- The choice of ranking measure is important
- Working with the joint posterior distribution over parameters allows us to identify more differences in rank
  - To be confirmed with simulation studies
- Hierarchical model building allows us to identify interesting patterns in the data and improve the ranking