Quiz 2
Empirical Bayes Estimation and Random Intercept Linear Models

From Tuesday's lecture on the two-stage model and empirical bayes estimation, answer the following questions:

1. In a two-stage model, the major sources of variation in an estimate of a regression parameter (e.g. log relative risk) are (check all that apply):
(a). statistical error the arises from imprecision in the finite set of measurements
(b). Bayesian error
(c). conjugate distribution error
(d). natural variation in the true parameter values
(e). stochastic correspondence deviations

2. In estimating the average parameter value (here, log relative risk) across cities, we should weight the city-specific estimates: (choose best answer):
(a). inversely proportional to the standard error
(b). proportional to the standard error
(c). inversely proportional to the statistical variance
(d). proportional to the statistical variance
(e). inversely proportional to the sum of the statistical and natural variance

3. When the statistical variance is small relative to the natural variance, we estimate each city's parameter value by: (choose best answers):
(a). the un-weighted average of all the city-specific estimates
(b). that city's maximum likelihood estimate
(c). the weighted average of all the city-specific estimates (d). a linear combination of the city-specific mle and the overall un-weighted average
(e). a linear combination of the city-specific mle and the overall weighted average

4. Relative to the mle, the empirical Bayes estimate for a city's parameter (e.g. log relative risk) is: (check all correct answers)
(a). is shrunk toward the overall estimate
(b). is more biased
(c). is more precise
(d). is less biased
(e). is less precise

From Wednesday's lecture on linear random intercept models, answer the following questions:

5. In the linear random intercept model example from the lecture, we define two sources of variation from measurements of the guinea pigs. These are:
(a). random variation in the outcome of interest measured within the same guinea pig over time
(b). measurement error in any weight measurement
(c). natural heterogeneity in the guinea pigs which may represent genetic variation
(d). statistical variation in the measurement of the rate of change of weight with time

6. When you specify a linear random intercept model, what type of correlation structure are you defining?
(a). an independence structure, i.e. no correlation
(b). an auto-regressive correlation structure, observations within units become less correlated over time
(c). an exchangeable correlation structure, the correlation is the same between any two measurements from the same unit (time is exchangeable)
(d). no correlation structure, we are estimating the ratio of natural variance to total variance.

7. Consider the following design: You have a random sample of 50 hospitals (indexed by i) and within each hospital you sample a varying number of surgery patients (indexed by j = 1, 2, …, n_i). The goal of the study is to model the average LOS as a function of a patient severity score. Suppose for now that we can model LOS as a linear variable (although it is really a count!). You specify the following model:

$$E(LOS_{ij}) = b0 + b1*severity_{ij} + u_i + e_{ij}$$

where $u_i \sim Normal(0, tau^2)$ and $e_{ij} \sim Normal(0, sigma^2)$.

In words, interpret tau^2 and sigma^2.

This is a linear random intercept model, which induces an exchangeable correlation structure among patients from the same hospital. In words explain what this means in this setting.