PhD Prelim Exam
METHODS

Summer 2010
(Given on 7/6/10)

Part 1.

A full-rank multiple regression model, $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \epsilon_i$, with three nonstochastic predictor variables (x_1, x_2, and the product $x_1 x_2$) was fit to a sample of size $n = 200$. You may assume that all standard multiple regression modeling assumptions are satisfied. Estimated coefficients and the estimated covariance matrix are below.

$$\begin{aligned} \widehat{\beta}_0 &= 0.0 \\ \widehat{\beta}_1 &= 0.2 \\ \widehat{\beta}_2 &= 1.3 \\ \widehat{\beta}_3 &= 0.1 \end{aligned} \qquad \widehat{\Sigma}_\beta = \begin{pmatrix} 0.06 & 0.03 & 0.02 & 0.05 \\ 0.03 & 0.08 & 0.02 & 0.01 \\ 0.02 & 0.02 & 0.10 & 0.02 \\ 0.05 & 0.01 & 0.02 & 0.05 \end{pmatrix}$$

1. Compute a two-sided 95% confidence interval for $\beta_1 - \beta_2$.
2. What are the advantages of reporting a confidence interval for $\beta_1 - \beta_2$ as compared to reporting a p-value and test statistic for the null hypothesis $H_0 : \beta_1 = \beta_2$ vs. $H_A : \beta_1 \neq \beta_2$?
3. What is the estimated slope for x_1 at $x_2 = 20$?
4. Compute a two-sided 95% confidence interval for the population slope for x_1 at $x_2 = 20$.
5. Compute a 95% prediction interval for a single future observation at $x_1 = 15$ and $x_2 = 10$. Since you do not have enough information to calculate the appropriate standard error, give its mathematical expression instead.

Part 2.

The sample means, variances, and sample sizes are given below for data from a completely randomized 2×2 factorial design. (The sample sizes are unequal due to a random loss of data.)

Factor A Level	Factor B Level	Mean	Variance	Sample Size
1	1	10.5	3.4	10
1	2	10.1	3.2	9
2	1	12.0	3.5	10
2	2	8.6	3.1	10

6. Express the following effects in terms of the unknown population means μ_{ij}, where μ_{ij} is the unknown population mean under level i of Factor A and level j of Factor B.

   (a) the main effect of Factor A

   (b) the $A \times B$ interaction effect

   (c) a simple effect of A at level 1 of B

7. Assuming equal variances (homoscedasticity), compute a two-sided 95% confidence interval for the $A \times B$ interaction effect.

8. Explain how you would construct a two-sided 95% confidence interval for the $A \times B$ interaction effect without assuming homoscedasticity.

Part 3.

A sample of $n = 10$ middle-aged males were each given 20 milligrams (mg) of a cholesterol-lowering drug. Their total cholesterol levels were measured once each week for four weeks. The corresponding sample means for Weeks 1-4 are

$$\bar{\mathbf{y}} = [240, 232, 226, 218]' .$$

The sample variance/covariance matrix for Weeks 1-4 is

$$\widehat{\Sigma} = \begin{pmatrix} 10 & 6 & 4 & 2 \\ 6 & 11 & 6 & 3 \\ 4 & 6 & 10 & 5 \\ 2 & 3 & 5 & 11 \end{pmatrix} .$$

9. The "population slope" of a line relating the population mean (μ_j) to Week (j) might be defined as $\sum_{j=1}^{4} c_j\mu_j$ for $c_1 = -3/20, c_2 = -1/20, c_3 = 1/20, c_4 = 3/20$. Compute a two-sided 95% confidence interval for $\sum_{j=1}^{4} c_j\mu_j$ and interpret the result.

10. Another way to analyze the slope in this context is to use a mixed model with Week as a predictor variable and a first-order autoregressive (AR(1)) covariance structure. Discuss the advantages and disadvantages of the mixed model approach versus the approach of Question 9.

Part 4.

An approximate two-sided confidence interval with endpoints of the form

$$\widehat{\theta} \pm z_{1-\alpha/2} \left[\widehat{\text{var}}(\widehat{\theta})\right]^{\frac{1}{2}} \ ,$$

where $\widehat{\theta}$ is assumed to be asymptotically normal, is sometimes called a "Wald" confidence interval. Further, an approximate two-sided confidence interval of the form

$$\left(g^{-1}\left(g(\widehat{\theta}) - z_{1-\alpha/2}\left[\widehat{\text{var}}(\widehat{g(\theta)})\right]^{\frac{1}{2}}\right), g^{-1}\left(g(\widehat{\theta}) + z_{1-\alpha/2}\left[\widehat{\text{var}}(\widehat{g(\theta)})\right]^{\frac{1}{2}}\right)\right)$$

is sometimes called a "transformed Wald" confidence interval.

Let Y_1 and Y_2 be independent random variables with $Y_1 \sim \text{Binomial}(n_1, \pi_1)$ and $Y_2 \sim \text{Binomial}(n_2, \pi_2)$.

11. Assuming the asymptotic normality of Y_1 and Y_2, derive a two-sided $100(1-\alpha)\%$ Wald confidence interval for $\theta = \pi_1 - \pi_2$.

12. Assuming the asymptotic normality of Y_1 and Y_2, derive a two-sided $100(1-\alpha)\%$ transformed Wald confidence interval for $\theta = \pi_1/\pi_2$ using $g(\theta) = \ln(\theta)$.

1. The estimated variance of $\widehat{\beta}_1 - \widehat{\beta}_2$ is $0.08 + 0.10 - 2(0.02) = 0.14$. The 95% CI $= 0.2 - 1.3 \pm 1.96\sqrt{0.14} = (-1.83, -0.37)$.

2. The CI provides information about both direction and the magnitude of the difference. The test only provides information about the direction of the difference.

3. $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 = \beta_0 + (\beta_1 + \beta_3 x_2)x_1 + \beta_2 x_2$, where $(\beta_1 + \beta_3 x_2)$ is the population "simple slope" for x_1. A point estimate of $\beta_1 + \beta_3 x_2$ is $\widehat{\beta}_1 + \widehat{\beta}_3 x_2 = 0.2 + 0.1(20) = 2.2$.

4. The estimated variance of $\widehat{\beta}_1 + \widehat{\beta}_3 x_2$ at $x_2 = 20$ is $0.08 + 400(0.05) + 2(20)(0.01) = 20.48$. The 95% CI is $2.2 \pm 1.96\sqrt{20.48} = (-6.67, 11.07)$.

5. $\widehat{y} = 0.0 + 0.2(15) + 1.3(10) + 0.1(15)(10) = 43.20$. $t_{0.975,196} = 1.972$. Let $X_h' = (1, 15, 10, 150)$. The standard error, $se_{\widehat{y}}$, is

$$\sqrt{\text{MSE}[1 + X_h'(X'X)^{-1}X_h]}\,.$$

The 95% prediction interval is $43.20 \pm 1.972 se_{\widehat{y}}$.

6. Express the following effects in terms of the unknown population means
   (a) main effect of A: $(\mu_{11} + \mu_{12})/2 - (\mu_{21} + \mu_{22})/2$
   (b) interaction effect: $\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}$
   (c) simple main effect of A at level 1 of B: $\mu_{11} - \mu_{21}$

7. Assuming equal variances (homoscedasticity), compute a two-sided 95% confidence interval for the $A \times B$ interaction effect.
   - pooled variance estimate $= \widehat{\sigma}^2 = (9(3.4)+8(3.2)+9(3.5)+9(3.1))/35 = 3.3$.
   - From the table, $t_{1-\alpha/2,df} = 2.03$ for $df = 35$ and $\alpha = 0.05$.
   - equal variance standard error $= \sqrt{\widehat{\sigma}^2 \sum c_j^2/n_j} = \sqrt{3.3(0.411)} = 1.16$
   - point estimate $= 10.5 - 10.1 - 12.0 + 8.6 = -3.0$ The 95% CI $= -3.0 \pm 2.03(1.16) = (-5.36, -0.64)$

8. Use a separate-variance standard error, $\sqrt{\sum c_j^2 \widehat{\sigma}_j^2 / n_j}$, and a Satterthwaite adjustment to the degrees of freedom.

9. Let $\mathbf{c}' = [-0.15, -0.05, 0.05, 0.15]$, $\widehat{\mu}' = [240, 232, 226, 218]$, and

$$\widehat{\Sigma} = \begin{pmatrix} 10 & 6 & 4 & 2 \\ 6 & 11 & 6 & 3 \\ 4 & 6 & 10 & 5 \\ 2 & 3 & 5 & 11 \end{pmatrix}$$

$$\begin{aligned} \mathbf{c}'\widehat{\mu} &= -3.6 \\ \mathbf{c}'\widehat{\Sigma}\mathbf{c} &= 0.465 \\ t_{0.975,9} &= 2.26 \end{aligned}$$

The 95% CI is $-3.6 \pm 2.26\sqrt{0.465} = (-5.14, -2.06)$. Interpretation: If all middle-aged men from which the sample was taken were given 20mg of the drug, the researcher can be 95% confident that the population mean cholesterol level would drop 2.06 to 5.14 points per week during weeks 1-4 (assuming linearity in this time frame).

10. Using the linear contrast approach, the only assumptions are independence among the 10 "subjects" and normality of the linear contrast. If these assumptions are satisfied, this approach provides an exact small-sample confidence interval and test. This approach is robust to moderate violations of the normality assumption and its robustness increases with larger sample sizes. However, missing data is a problem with the linear contrast approach because a subject is dropped unless all four measurements (weeks 1-4) have been obtained for that subject. With the mixed model approach, all of the available data may be used when there is missing data, and the mixed model approach can accommodate covariates. However, the tests and confidence intervals for the mixed model are large-sample procedures and may not perform well for a sample of $n = 10$. Furthermore, misspecification of the residual covariance structure (e.g., AR(1)) introduces a bias in the standard errors.

11. Assuming the asymptotic normality of Y_1 and Y_2, derive a two-sided $100(1-\alpha)\%$ Wald confidence interval for $\theta = \pi_1 - \pi_2$.

$$\widehat{\pi}_1 = Y_1/n_1$$

$$
\begin{aligned}
\widehat{\pi}_2 &= Y_2/n_2 \\
\widehat{\text{var}}(\widehat{\pi}_1) &= \widehat{\pi}_1(1-\widehat{\pi}_1)/n_1 \\
\widehat{\text{var}}(\widehat{\pi}_2) &= \widehat{\pi}_2(1-\widehat{\pi}_2)/n_2 \\
\widehat{\theta} &= \widehat{\pi}_1 - \widehat{\pi}_2 \\
\widehat{\text{var}}(\widehat{\theta}) &= \widehat{\pi}_1(1-\widehat{\pi}_1)/n_1 + \widehat{\pi}_2(1-\widehat{\pi}_2)/n_2
\end{aligned}
$$

The $100(1-\alpha)\%$ confidence interval for $\pi_1 - \pi_2$ is

$$\widehat{\pi}_1 - \widehat{\pi}_2 \pm z_{1-\alpha/2}[\widehat{\pi}_1(1-\widehat{\pi}_1)/n_1 + \widehat{\pi}_2(1-\widehat{\pi}_2)/n_2]^{\frac{1}{2}}$$

12. Assuming the asymptotic normality of Y_1 and Y_2, derive a two-sided $100(1-\alpha)\%$ transformed Wald confidence interval for $\theta = \pi_1/\pi_2$ using $g(\theta) = \ln(\theta)$.

$$
\begin{aligned}
\widehat{\theta} &= \widehat{\pi}_1/\widehat{\pi}_2 \\
\partial \ln(\theta)/\partial \pi_1 &= 1/\pi_1 \\
\partial \ln(\theta)/\partial \pi_2 &= -1/\pi_2
\end{aligned}
$$

Using the "delta" method, $\widehat{\text{var}}[\ln(\widehat{\theta})] \approx (1-\widehat{\pi}_1)/(\widehat{\pi}_1 n_1) + (1-\widehat{\pi}_2)/(\widehat{\pi}_2 n_2)$. A $100(1-\alpha)\%$ CI for π_1/π_2:

$$\exp\{\ln(\widehat{\pi}_1/\widehat{\pi}_2) \pm [(1-\widehat{\pi}_1)/(\widehat{\pi}_1 n_1) + (1-\widehat{\pi}_2)/(\widehat{\pi}_2 n_2)]^{\frac{1}{2}}\}$$

A longitudinal study of child development, initiated by the Harvard School of Public Health, provided data on the heights of a sample of 67 girls who were 7 years old in 1929. The height of each girl (in cm) was measured annually until each girl reached 18 years of age. Consequently, 12 height measurements were recorded on each girl in the sample. You may consider this sample of girls to be a simple random sample from the population of 7 year old girls in the Boston metropolitan area in 1929.

Five graphical displays of the data are shown below. The box plots shown in Figure 1 are *schematic plots* with the whiskers drawn from the edges of the boxes to either the most extreme observation or the end of the inner fence, whichever is closer to the median. (An inner fence starts at a quartile and extends 1.5 times the inter-quartile range beyond the quartile.) Observations beyond the inner fence are displayed as open squares.

Figure 1: Box Plots of Heights of 67 Girls

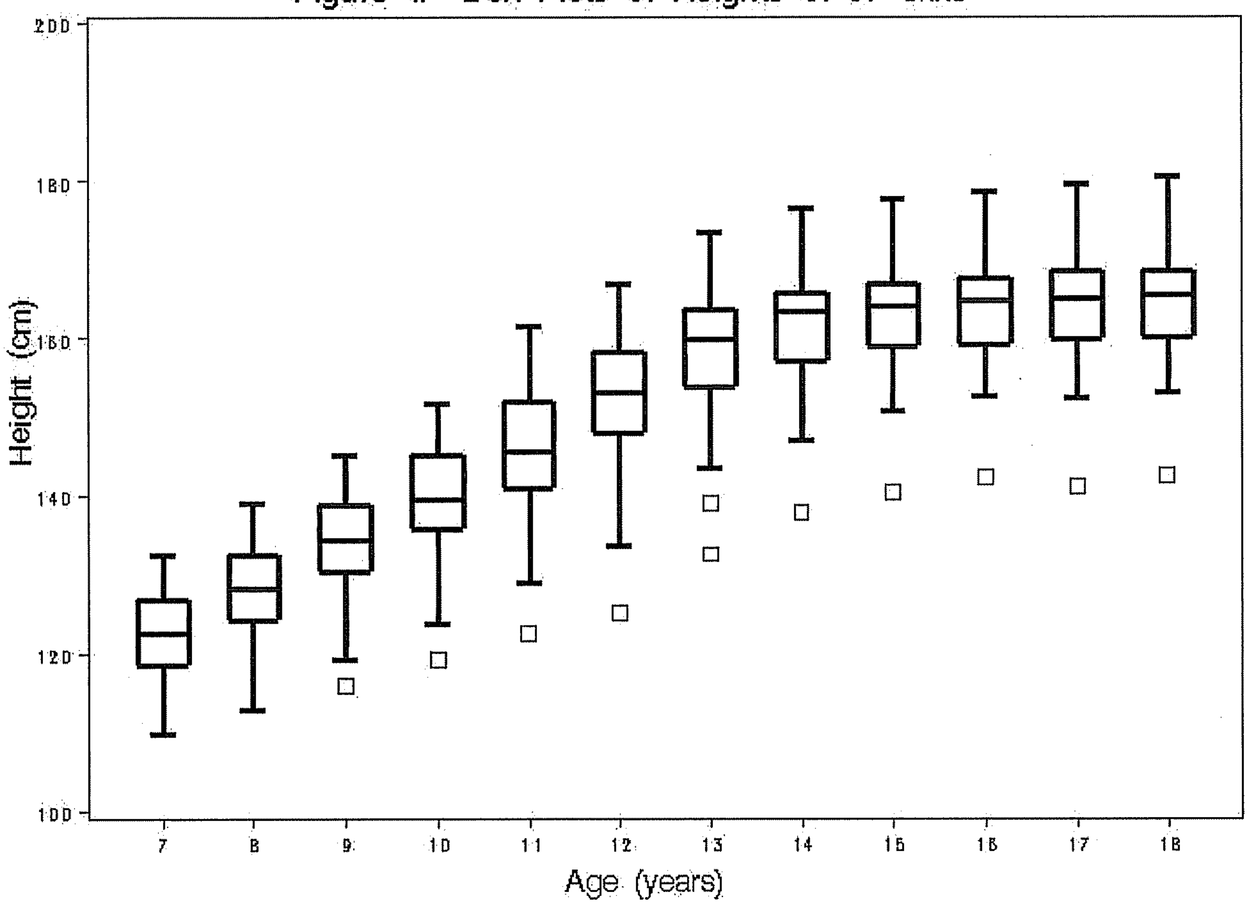


Figure 2 shows the individual growth profiles for the 67 girls in the study. For an individual girl, the growth profile is constructed by connecting the girl's heights at consecutive ages with straight line segments.

Figure 2: Heights of 67 Girls

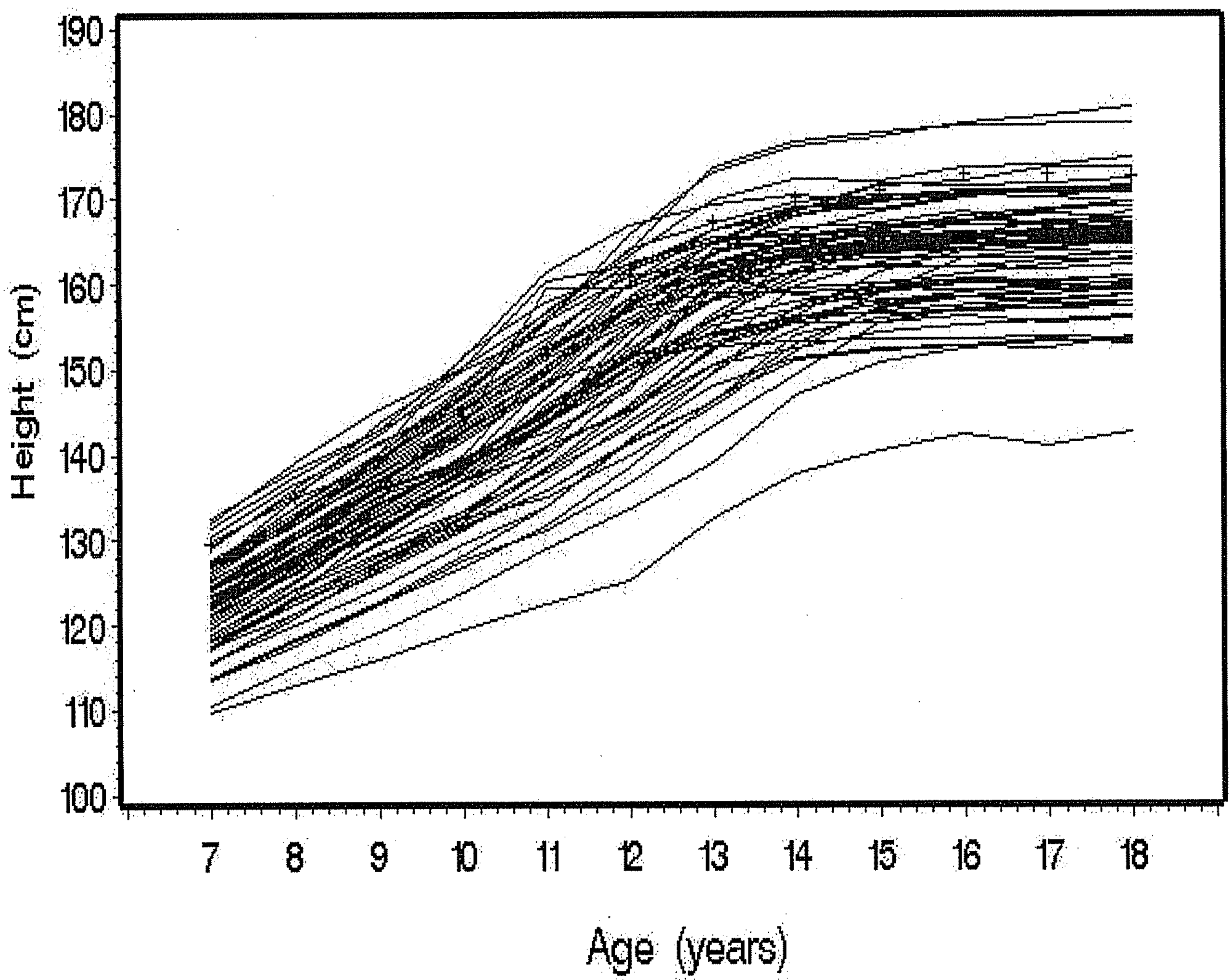


Figure 3 shows the estimates of overall lagged correlations for the 67 girls in the study. To estimate the correlation with a lag of 9 years, for example, any pair of observations taken 9 years apart on the same girl are included in the calculation. There would be three pairs of observations of each girl in this case: heights measured at ages 7 and 16 years, heights measured at ages 8 and 17 years, and heights measured at ages 9 and 18 years.

Figure 3: Estimated Lagged Correlations

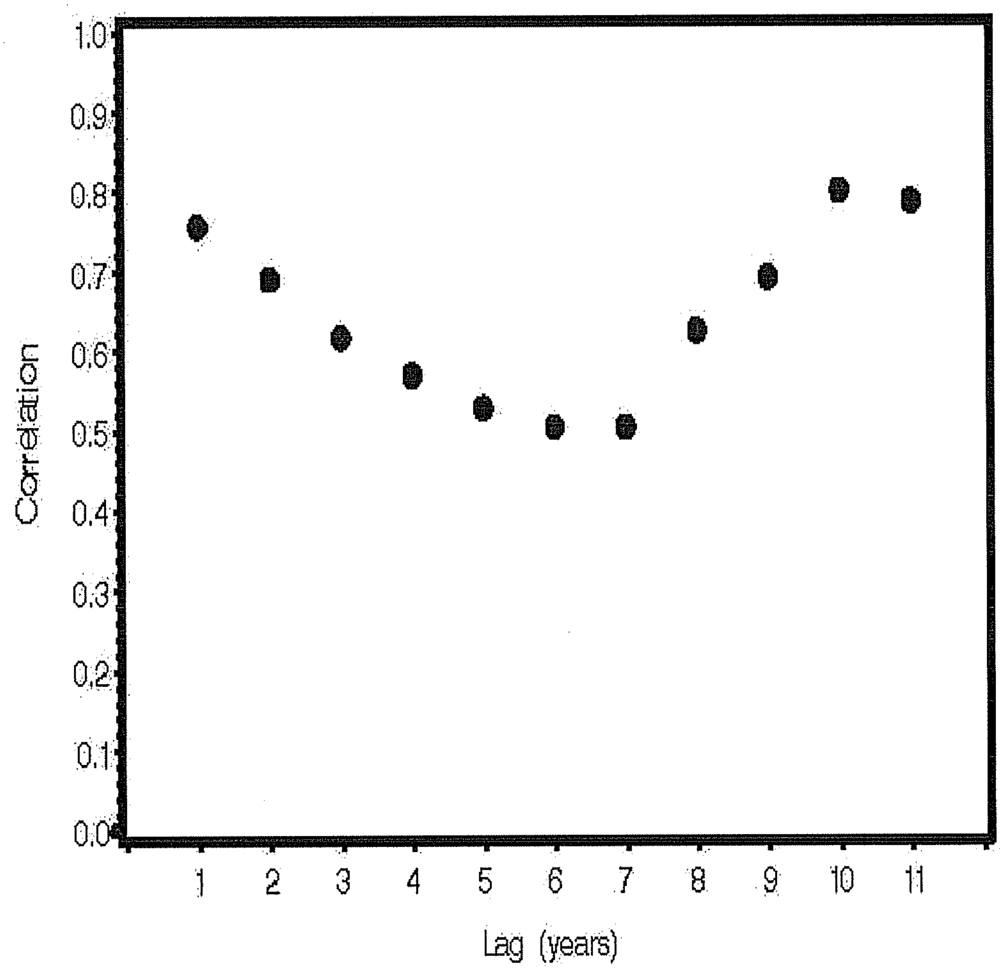
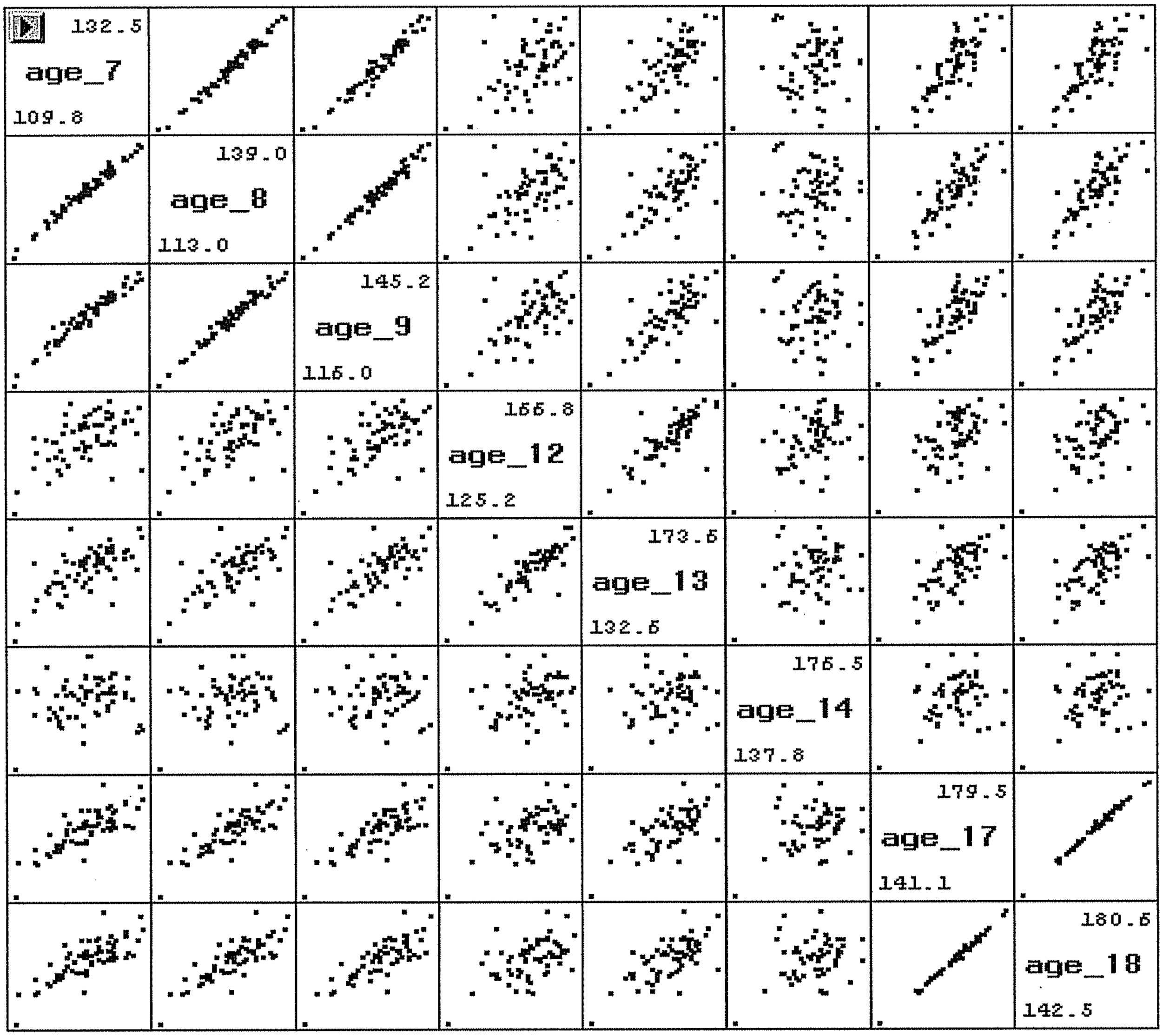


Figure 4 contains a scatterplot matrix that displays relationships among heights at ages 7, 8, 9, 12, 13, 14, 17 and 18 years. In order to keep the number of plots in the scatterplot matrix small enough for easy viewing, heights at ages 10, 11, 15 and 16 are not included.

Figure 4: Scatter Plot Matrix



(a) What do these plots suggest about features that should be incorporated into a model for describing the changes in heights of girls between ages 7 and 18?

Six models of the following form were fit to the data:

$$Y_{ij} = \mu_j + \epsilon_{ij}$$

where

$$\begin{aligned} Y_{ij} &= \text{height of the } i\text{-th girl at the } j\text{-th time point } (j = 1 \text{ for 7 years of age,} \\ &\quad j = 2 \text{ for 8 years of age, etc.)} \\ \mu_j &= \text{population mean height at the } j\text{-th time point} \\ \epsilon_{ij} &= \text{random error} \end{aligned}$$

Let $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{i,12})'$ denote the vector of random errors associated with the twelve height measurements taken on on the i-th child. Each model assumes that the ϵ_i's are mutually independent random vectors and each ϵ_i has a multivariate normal distribution with a zero mean vector and some covariance matrix. The form for the covariance matrix differs across the models.

Model 1: (Gauss-Markov model)

$$\text{Var}(\epsilon_{ij}) = \sigma_\epsilon^2 \text{ and } \text{Cov}(\epsilon_{ij}, \epsilon_{ik}) = 0 \text{ for all } j \neq k.$$

Model 2: (AR(1) model with homogeneous variances)

$$\text{Var}(\epsilon_{ij}) = \sigma_\epsilon^2 \text{ and } \text{Cov}(\epsilon_{ij}, \epsilon_{ik}) = \rho^{|j-k|}\sigma_\epsilon^2 \text{ for all } j \neq k \text{ and some correlation } 0 < \rho < 1.$$

Model 3: (AR(1) model with heterogeneous variances)

$$\text{Var}(\epsilon_{ij}) = \sigma_{\epsilon,j}^2 \text{ and } \text{Cov}(\epsilon_{ij}, \epsilon_{ik}) = \rho^{|j-k|}\sigma_{\epsilon,j}\sigma_{\epsilon,k} \text{ for all } j \neq k \text{ and some correlation } 0 < \rho < 1.$$

Model 4: (Toeplitz model with homogeneous variances)

$$\text{Var}(\epsilon_{ij}) = \sigma_\epsilon^2 \text{ and } \text{Cov}(\epsilon_{ij}, \epsilon_{ik}) = \rho_{|j-k|}\sigma_\epsilon^2 \text{ for all } j \neq k, \text{ with correlations } \rho_1, \rho_2, \ldots, \rho_{11}.$$

Model 5: (Toeplitz model with heterogeneous variances)

$$\text{Var}(\epsilon_{ij}) = \sigma_{\epsilon,j}^2 \text{ and } \text{Cov}(\epsilon_{ij}, \epsilon_{ik}) = \rho_{|j-k|}\sigma_{\epsilon,j}\sigma_{\epsilon,k} \text{ for all } j \neq k, \text{ with correlations } \rho_1, \rho_2, \ldots, \rho_{11}.$$

Model 6: (Model with an arbitrary 12×12 covariance matrix for the 12 height measures taken on each girl)

$$\text{Var}(\epsilon_{ij}) = \sigma_{\epsilon,j}^2 \text{ and } \text{Cov}(\epsilon_{ij}, \epsilon_{ik}) = \sigma_{\epsilon,jk} \text{ for all } j \neq k.$$

For each of these models, the parameters are restricted to define a positive definite covariance matrix for ϵ_i. REML estimation was used to estimate the covariance matrix parameters. The values of the logarithm of the REML likelihood function and corresponding AIC and BIC values are presented in the following table:

Model	log(REML Likelihood)	AIC	BIC
1	−2637.7	5277.5	5279.6
2	−1584.4	3172.8	3178.2
3	−1463.2	2952.4	2981.4
4	−1541.4	3106.8	3133.2
5	−1328.2	2702.4	2753.1
6	−1287.8	2731.6	2893.5

(b) Which of these six models is best supported by the data? Support your decision with statistical reasoning.

A model might be considered to provide a more parsimonious description of changes in the mean heights across time. The model is

Model 7: $$Y_{ij} = \beta_0 + \beta_1 X_j + \beta_2 (X_j - 14) I_{(X_j \geq 14)} + \epsilon_{ij}$$

where

Y_{ij} = height of i-th girl at the j-th time point

$X_j = j + 6$ = age at the j-th time point, $j = 1, \ldots, 12$

$$I_{(X_j \geq 14)} = \begin{cases} 1 & \text{if } X_j \geq 14 \\ 0 & \text{otherwise} \end{cases}$$

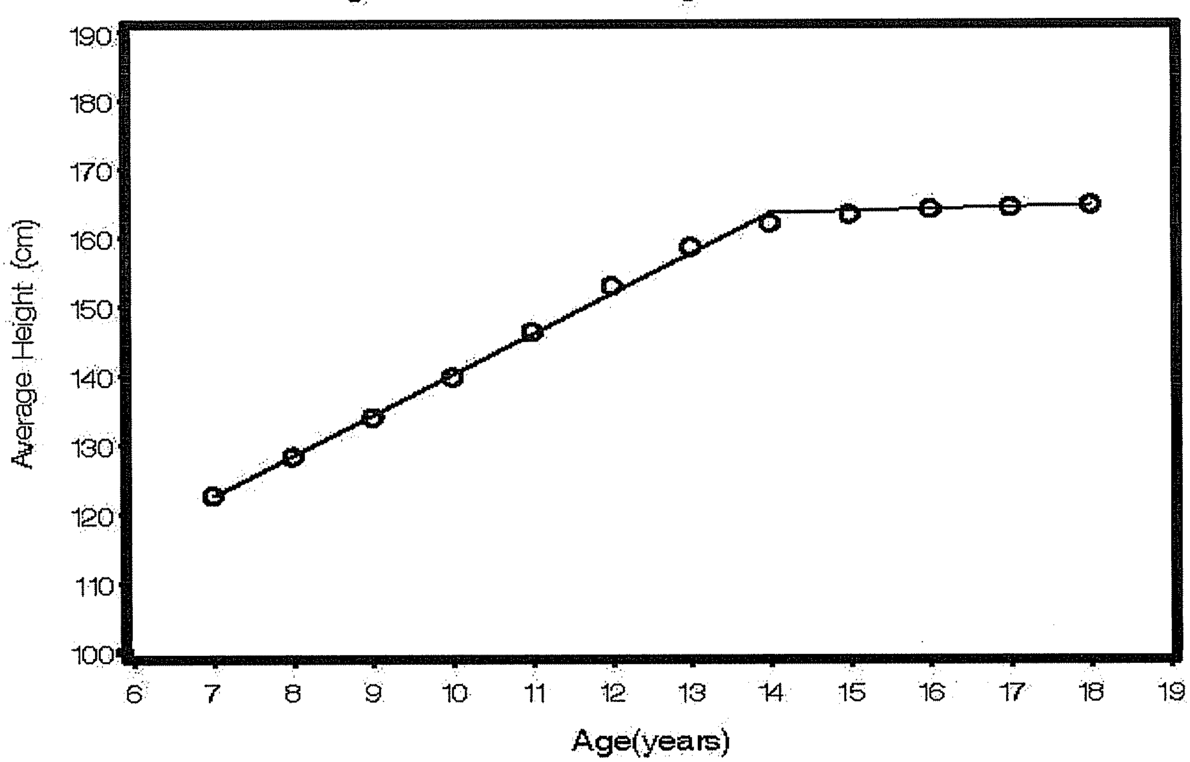
$\epsilon_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,12})'$ is a vector of random errors for the 12 height measurements obtained from the i-th girl .

The least squares estimate of the growth curve is:

$$\hat{Y} = 81.93 + 5.80x - 5.46(x - 14) I_{(x \geq 14)}.$$

Figure 5 shows the estimated growth curve as a pair of adjoining line segments. It also shows the sample means for the heights of the 67 girls for each year as open circles.

Figure 5: Predicted Heights of Girls



(c) Let $\hat{\beta}_{OLS} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)'$ denote the vector of ordinary least squares estimators for the regression coefficients. Assume the ϵ_i's are mutually independent multivariate normal random variables with a vector of zeros as the mean vector and an arbitrary positive definite and symmetric covariance matrix Σ. Note that Σ may not be a scalar multiple of the identity matrix. It could be one of the other covariance models considered in Models 2 through 6. In this situation, what is the distribution of $\hat{\beta}_{OLS}$?

(d) Considering the distribution you derived in part (c), how would you estimate the covariance matrix for $\hat{\beta}_{OLS}$.

(e) Assuming Σ is known, give a formula for $\hat{\beta}_{GLS}$ the generalized least squares estimator of β, and compare the distribution of $\hat{\beta}_{GLS}$ to the distribution of $\hat{\beta}_{OLS}$. As part of your comparison, address the issue of which estimator is more efficient.

Consider Model 7 with the additional assumption that the vector of random errors ϵ_i, associated with the set of 12 height measurements taken on the *i-th* girl in the study, is a multivariate normal random variable with a zero mean vector and an arbitrary positive definite and symmetric 12×12 covariance matrix Σ. Also assume that the ϵ_i's are mutually independent. Results from REML estimation, using the MIXED procedure in SAS, are shown below. Satterthwaite degrees of freedom were used to obtain p-values for approximate t-tests.

Solution for Fixed Effects

Effect	Estimated Coefficient	Standard Error	DF	t Value	Pr > |t|
Intercept	83.7679	0.6935	65.6	120.79	<.0001
X	5.5637	0.0622	65.9	89.43	<.0001
X*I(X>=14)	-4.9675	0.0715	66.3	-69.52	<.0001

Covariance Matrix for Fixed Effects

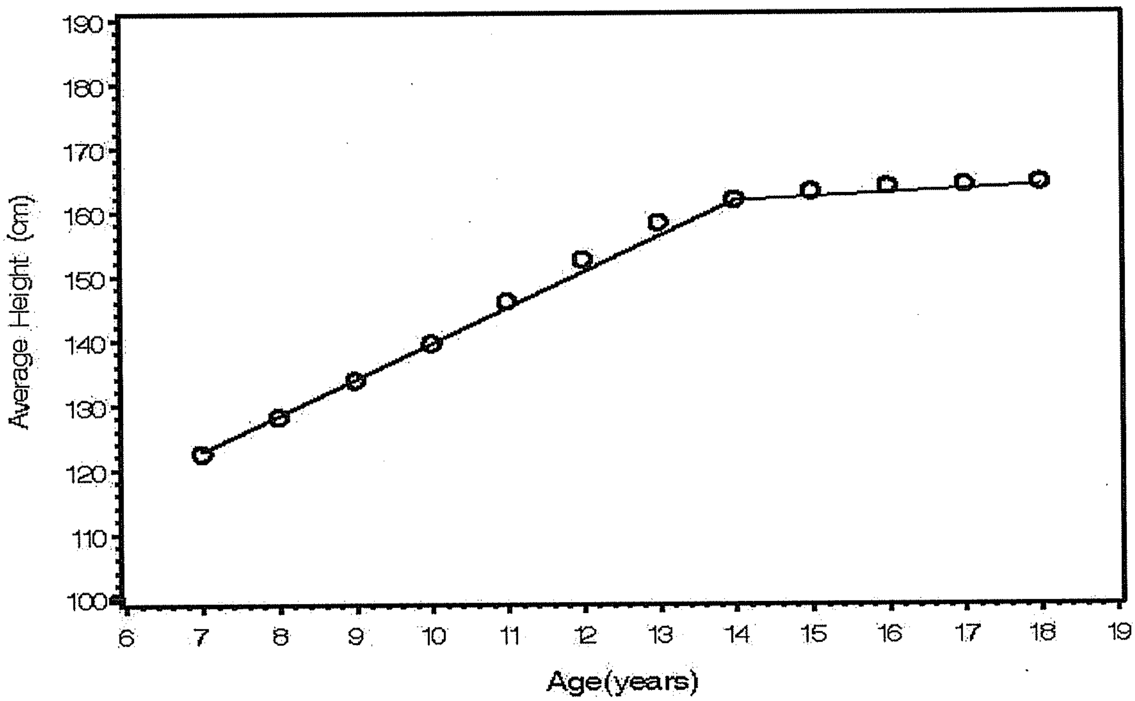
Effect	Intercept	X	X*I(X>=14)
Intercept	0.48090	-0.02035	0.01946
X	-0.02035	0.00387	-0.00364
X*I(X>=14)	0.01946	-0.00364	0.00511

Fit Statistics:

-2 REML Log Likelihood	2695.0
AIC (smaller is better)	2851.0
BIC (smaller is better)	3023.0

Figure 6 shows the estimated growth curve (using REML estimation to estimate Σ) as a pair of adjoining line segments. It also shows the sample means for the heights of the 67 girls for each year as open circles.

Figure 6: Predicted Heights of Girls



(f) Briefly describe how the likelihood function used in REML estimation of the covariance parameters is obtained. You may use matrix notation and other standard statistics notation, but you do not have to provide an explicit formula for the likelihood function.

(g) The estimator for β provided by the REML option in the MIXED procedure in SAS (call it $\hat{\beta}_{GLSREML}$) is evaluated using the formula you gave for $\hat{\beta}_{GLS}$ in part (e), with REML estimates substituted the parameters in Σ. Describe how the distribution of $\hat{\beta}_{GLSREML}$ compares to the distribution of $\hat{\beta}_{GLS}$? Which properties do they have in common? How do they differ?

(h) Use the estimates of the parameters and the estimate of the covariance matrix for the parameters shown above to estimate the average annual rate of growth (slope of the line) for girls in the Boston Metropolitan Area at 16 years of age. Also, construct an approximate 95% confidence interval for the average annual rate of growth (slope of the line) for girls in this population at 16 years of age.

(i) Explain how you would test the null hypothesis that Model 7 provides an accurate description of how mean heights for the population of girls in the Boston Metropolitan Area change between 7 and 18 years of age, i.e., explain how you would test $H_o : E(Y_{ij}) = \beta_0 + \beta_1 X_j + \beta_2(X_j - 14)I_{(X_j \geq 14)}$ for all $j = 1, 2, ..., 12$. The alternative is $H_a : E(Y_{ij}) \neq \beta_0 + \beta_1 X_j + \beta_2(X_j - 14)I_{(X_j \geq 14)}$ for at least one value of j. This null hypothesis does not involve Σ, but your test procedure should not ignore the possibility of heterogeneous variances at different ages or the possibility of correlations among repeated measures of heights on individual girls.

(j) Figures 1 and 2 indicate that the shortest girl at 7 years of age remained the shortest girl through 18 years of age. By age 9, her height was far enough below the height of the other girls in the sample that she was outside the inner fences of the box plots shown in Figure 1. The growth profile for this girl, displayed in Figure 2, is consistently below the growth profile for any other girl in the sample. Explain how you would perform a test to determine if the growth profile for this girl is an outlier relative to the model used in the REML estimation, i.e., Model 7 with the additional assumptions that the vectors of random errors associated with profiles for different girls are mutually independent normal random vectors with zero mean vector and covariance matrix Σ. Give a description of your testing procedure and indicate how you would reach a conclusion.

(a) The box plots show that the median (mean) heights increase in roughly a straight line trend from ages 7 through 14 and then growth slows down and median (mean) heights approach an asymptote for young adult height. There also appears to be more variation in heights for ages 11, 12 and 13 than for other heights, suggesting that a model with heterogeneous variances across ages may be needed. The sample profiles displayed in Figure 2 indicate that repeated measurements of heights on the same girl are positively correlated, because profiles for some girls clearly stay below the median heights and profiles for other girls clearly stay above median heights. Figures 1 and 2 also indicate that the shortest girl at age 7 has a growth profile well below the growth profiles of the other girls in the study. Figure 3 indicates that correlations among repeated measures of heights are all positive, but correlations become weaker as the time lag increases from one through 7 years. Correlations become stronger for time lags longer than 7 years. The scatterplot matrix in figure 4 reveals strong positive correlations in heights at ages 7, 8 and 9 year and strong positive correlation between the heights at ages 17 and 18. Correlations between heights at ages 12, 13, and 14, when girls are growing rapidly, are weaker and there is more variability in heights at those ages. Correlations between heights during this fast growth period with heights at either younger or older ages are the weakest. This helps to explain the trend in the lagged correlations in figure 3, but the scatterplot provides more information than the lagged correlation plot. The scatterplot also shows that correlations among heights at the three early ages and correlations among the two oldest ages are stronger than figure 3 indicates. One would not expect a compound symmetry model or a simple autoregressive model to provide an adequate description of the covariance structure for repeated height measurements taken on the same girl.

(b) These six models fit the most general model to the mean heights by simply allowing for a different mean height for each age. The comparison of these models focuses on the determination of a covariance model for the 12 repeated measurements of heights taken on individual girls. Both the AIC and BIC criteria for penalized log-likelihoods are minimized for model 5, indicating that the Toeplitz model with heterogeneous variances provides a better estimate of the covariance matrix for the repeated measures than the first four models. The increased value of the REML log-likelihood indicates that the more general covariance structure provided by model 6 provides a better fit to the observed variance and covariances. The value of -2 times the difference in the REML log-likelihoods for models 5 and 6 is 80.8. Comparing this to the percentiles of a chi-square distribution with 78-23=55 degrees of freedom yields a p-value of 0.0133. The general covariance model may provide a slightly better description of the covariance pattern, but it appears that the Toeplitz model with heterogeneous variance is a good approximation. Either covariance model should lead to essentially the same inferences about mean heights.

(c) $\hat{\beta}_{OLS} \sim N(\beta, \sum_{i=1}^{67}(X_i'X_i)^{-1}X_i'\Sigma X_i(X_i'X_i)^{-1})$, where X_i are the rows of the model matrix for the i-th girl in the sample. The least squares estimator is unbiased and it has a multivariate normal distribution.

(d) A consistent estimate of $\text{Var}(\hat{\beta}_{OLS})$ is $\sum_{i=1}^{67}(X_i'X_i)^{-1}X_i'\hat{\Sigma}X_i(X_i'X_i)^{-1})$, where $\hat{\Sigma} = (64)^{-1}\sum_{i=1}^{67}(Y_i - \hat{Y})(Y_i - \hat{Y})'$, Y_i is the vector of observed heights for ages 7 through 18 for the i-th girl in the sample, and $\hat{Y}$ is the vector of estimated mean heights for ages 7 through 18 provide by Model 7.

(e) $\hat{\beta}_{GLS} \sim N(\beta, (\sum_{i=1}^{67} X_i'\Sigma^{-1}X_i)^{-1})$. The generalized least squares estimator of β is also unbiased and has a multivariate normal distribution, but it has a smaller covariance matrix than the ordinary least squares estimator in the sense that $\text{var}(a'\hat{\beta}_{GLS}a) \le \text{var}(a'\hat{\beta}_{OLS}a)$ for any non-random vector a.

(f) The likelihood function for a set of residuals is used to obtain the REML likelihood function. First, ordinary least squares estimation is used to obtain $\hat{\beta}_{OLS}$ for model 7. Denoting the 3×804 model matrix for the complete set of data for all 67 girls as X, and the corresponding vector of observed heights as Y, $\hat{\beta}_{OLS}=(X'X)^{-1}X'Y)$ and the vector of residuals is $e = Y - X\hat{\beta}_{OLS} = (I - X(X'X)^{-1}X')Y$. Since e must satisfy three linear constraints, e has a degenerate normal distribution. Using any 801×804 matrix M with full row rank, provides a multivariate normal random vector Me with a zero mean vector and a non-singular covariance matrix $M(I - X(X'X)^{-1}X')\Delta_\Sigma(I - X(X'X)^{-1}X')M'$, where Δ_Σ is a block diagonal matrix with Σ repeated 67 times as the diagonal block. The multivariate normal distribution for Me does not depend on β and it log-likelihood function, called the REML log-likelihood function, is maximized with respect to the parameters for Σ to obtain the REML estimates.

(g) $\hat{\beta}_{GLSREML}$ is not a linear function of the observed responses Y. It is not an unbiased estimator and it does not exactly have a multivariate normal distribution. It does have an asymptotic normal distribution that approaches the distribution for $\hat{\beta}_{GLS}$ as the number of girls in the sample increases.

(h) $\hat{\text{Slope}} = 5.5637 + (-4.96571) = 0.5962$ with standard error computed as $\sqrt{.00387 + .00511 + 2(-.00364)} = .041231$. An approximate 95 percent confidence interval is $0.5962 \pm (1.9967)(0.041231) \Rightarrow (0.514, 0.679)$.

(i) Students could discuss the construction of a likelihood ratio test, a Wald-type test, or a score test. Comparing REML log-likelihoods for models 6 and 7 is not a correct approach.

(j) Arrange the data so that the vector of height observations for the shortest girl is Y_{67}. One approach is to delete the vector of height observations on the shortest girl from the data and re-fit model 7 using only the data from the remaining 66 girls to obtain a new estimate of the regression coefficients, denoted by $\hat{\beta}_{GLSREML-67}$, and a new REML estimate of the covariance matrix for ϵ_i, denoted by $\hat{\Sigma}_{-67}$. Then, compute a Wald statistic $X_{67}^2 = (Y_{67} - X_{67}\hat{\beta}_{GLSREML-67})(\sum_{i=1}^{66} X_i'\hat{\Sigma}_{-67}^{-1}X_i)^{-1}(Y_{67} - X_{67}\hat{\beta}_{GLSREML-67})'$ and reject the null hypothesis that the growth profile for the smallest girl is not an outlier if X_{67}^2 is too large. Since the growth profile for the smallest girl was selected, this statistic will not have a chi-squared distribution when the null hypothesis is true. Students can describe a method for simulating the null distribution of this test statistic.

1 Problem Background

In Iowa, property taxes are used to support public services in the state. The largest single use of revenue generated by property taxes is K-12 public education. Property is defined primarily as what is called "real property," meaning land and buildings. Such properties are categorized into groups such as residential, commercial, agricultural, industrial, and a few others. The two largest categories by revenue are residential (producing about 47% of property taxes collected) and commercial (producing about 29% of taxes collected) and we will focus on commercial properties.

Although the "tax rate," which is taxable amount per dollar value of property, is determined through a rather involved process, once this rate is determined the amount of taxes levied becomes a simple function of the value of the property. Property values are assessed by elected public officials called assessors. There is one assessor for each of the 99 counties in Iowa, plus an additional 8 assessors for particular cities. The region an assessor is responsible for is called an "assessment jurisdiction" and so there are 107 jurisdictions in the state.

The state Department of Revenue is responsible for "equalizing" property tax rates for each assessment jurisdiction every two years. It does this by comparing assessed values to sales figures for properties that have been sold in the most recent year. The current procedure is as follows. For each assessment jurisdiction, the Department of Revenue computes the ratio of assessed value to selling price for each property sold as $y_{i,j} =$ (assessed value / sale price) for property j in jurisdiction i. The median of this ratio is computed for each jurisdiction which we will call M_i. If $M_i < 0.95$ or $M_i > 1.05$ then all of the assessed values in that jurisdiction and category are decreased or increased to make the median equal to 1.0.

The Department of Revenue is concerned that this may not be the best procedure to use for the process of equalization, and they are seeking advice from the Department of Statistics at Iowa State. In particular, the Department of Revenue is concerned that the empirical distribution of ratios across all jurisdictions usually appears to be skewed to the right. Also, the number of commercial properties sold in any given jurisdiction is fairly small, usually from single digits (e.g., 7) to the mid-teens (e.g., 16), although some of the larger jurisdictions have larger numbers of sales (e.g., 30). The task given to us is to model this situation on a state-wide basis, without focusing on a particular specific inference objective. That is, we are being asked to look at the problem from a statistical viewpoint first, in the hope that such consideration will lead to meaningful insights for the actual task of property tax equalization.

Assume that we have available to us data from each of the 107 jurisdictions with data records for individual properties containing a jurisdiction identifier, the assessed value and selling price, and the ratio of assessed value to sales price.

2 Modeling Ratios

We first consider how we might directly model ratios of assessed value to sales price in the collection of jurisdictions across the state. Figure 1 presents a histogram of the ratios for all properties sold in the entire state.

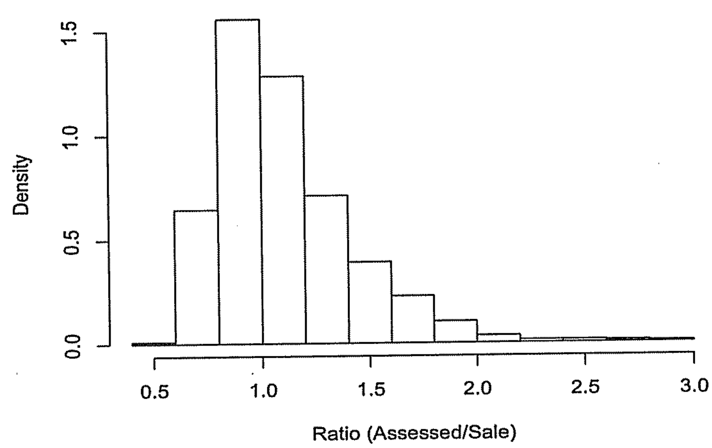


Figure 1: Marginal histogram of all ratios in the state.

Summary values corresponding to the histogram of Figure 1 are:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.5275	0.8792	1.0490	1.1060	1.2430	2.8990

Figure 2 presents histograms of the ratios for four particular jurisdictions individually and also a line plot of ratios within those jurisdictions. There are a total of 107 jurisdictions, but assume these histograms and line plot are representative of the types of empirical distributions that would be seen in the entire collection.

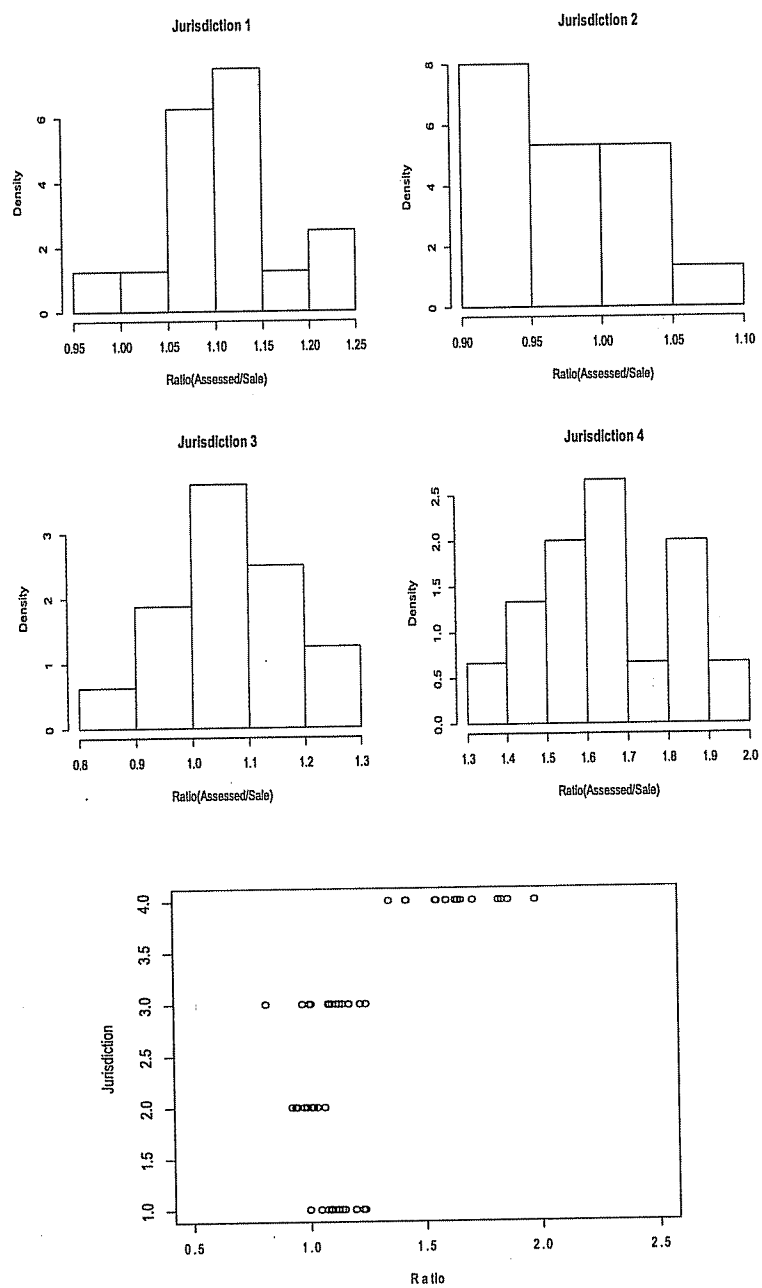


Figure 2: Histograms and a line plot of ratios in four individual jurisdictions. Sample sizes are $n = 15$ for jurisdiction 1, $n = 7$ for jurisdiction 2, $n = 16$ for jurisdiction 3 and $n = 15$ for jurisdiction 4

We may compute the means and standard deviations of the individual property ratios within each of the 107 jurisdictions. Figure 3 presents histograms of these quantities. Summary statistics for the means and standard deviations of ratios within jurisdictions are presented in Table 1.

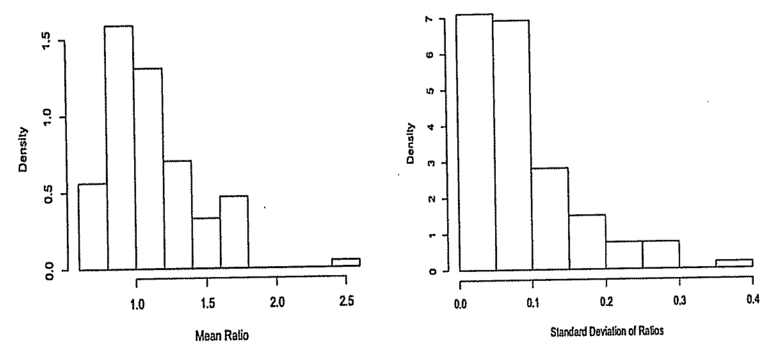


Figure 3: Histograms of mean and standard deviation of ratios in individual jurisdictions for all 107 jurisdictions in the state.

Quantity	Sample Size	Mean	Standard Deviation
Mean Ratio	107	1.1085	0.3057
Standard Deviation of Ratios	107	0.0893	0.0674

Table 1: Summary statistics for means and variance of ratios within jurisdictions.

Using the means and variances of ratios within jurisdictions we may also construct the graphs of Figure 4, which show the variances plotted against the means in the left panel and the log standard deviations plotted against the log means in the right panel. An ordinary least squares fit of log standard deviations to log means results in the equation $y = -2.798 + 2.071x$, where y is log standard deviation and x is log mean.

Let $R_{i,j}$ be a random variable connected with the ratio of assessed value to sale price for property j in jurisdiction i; $i = 1, \ldots K$, $j = 1, \ldots, n_i$. A possible model is that the $R_{i,j}$ are conditionally independent given the λ_i's, and λ_i for $i = 1, \ldots, K$ are independent and identically distributed as follows,

$$\begin{aligned} R_{i,j}|\lambda_i &\sim \text{ iid } N(\lambda_i,\, \gamma\lambda_i^{\theta}) \\ \log(\lambda_i) &\sim \text{ iid } N(\mu,\, \sigma^2) \end{aligned} \tag{1}$$

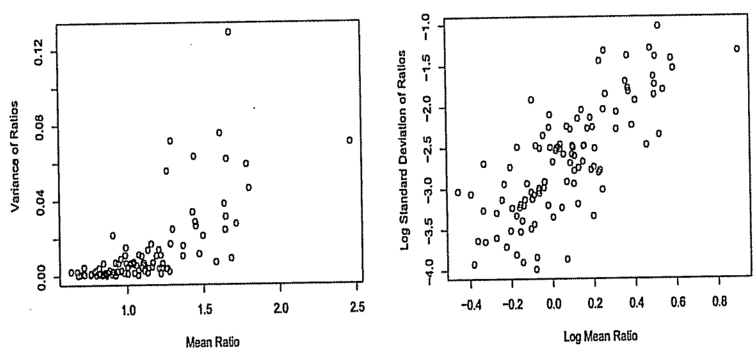


Figure 4: Means and variances of ratios (left) and log means and log standard deviations of ratios (right).

ANSWER QUESTIONS 1, 2, 3, 4 and 5 NOW

3 Modeling Assessed Values and Sale Prices

The ratios of Section 2 are constructed from the more fundamental quantities of assessed value and sale price for individual properties, so we might choose to examine those fundamental quantities in more detail. It would seem that, if the process of assessing property is meaningful, then more valuable properties should have higher assessed values than less valuable properties. Assuming that true value is reflected in sales price, there should then be a correlation between sales price and assessed value. Figure 5 presents the scatterplot of mean assessed values against mean sales prices for the assessment jurisdictions. It appears from this figure that average values of properties differ substantially among the various jurisdictions. This is confirmed by Figure 6 which shows line plots of individual assessed values in five jurisdictions.

We can also examine the relation between assessed values and sale prices within jurisdictions for at least a small number of jurisdictions at a time. Figure 7 presents plots of assessed values against sale prices for individual properties within two groups of four particular jurisdictions (note that the plotting symbols do not have the same meaning in the two panels of this figure – they just denote different jurisdictions in each plot).

One model that is suggested for this view of the problem is as follows. Let $Y_{i,j}$ be a random variable associated with the assessed value for the j^{th} property in the i^{th} jurisdic-

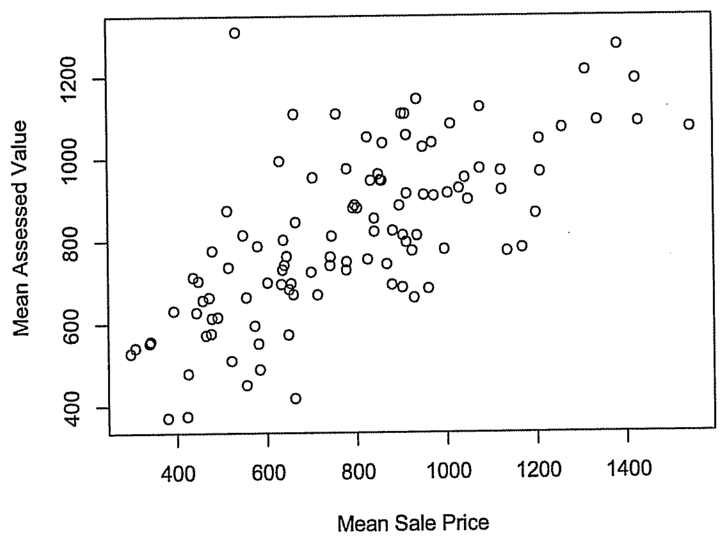


Figure 5: Mean sales prices and mean assessed values for the 107 assessment jurisdictions.

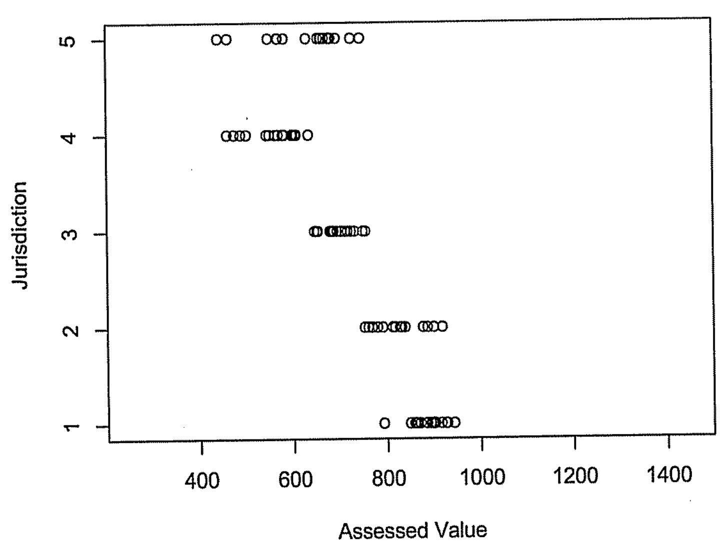


Figure 6: Line plots of assessed values within five particular individual jurisdictions.

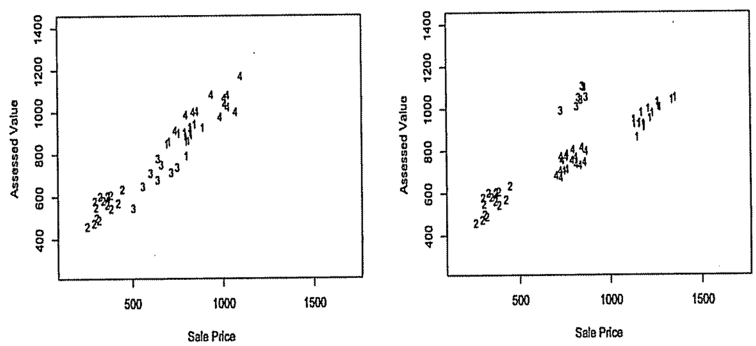


Figure 7: Assessed values versus sale prices for eight particular jurisdictions, four in the left panel and four in the right panel.

tion, $i = 1, \ldots, K$, $j = 1, \ldots, n_i$. Let $x_{i,j}$ denote the sale price of that same property, and consider these values as known, rather than the realization of a random process. Consider the model

$$Y_{i,j} = \beta x_{i,j} + \sigma \epsilon_{i,j}, \tag{2}$$

where the $\epsilon_{i,j}$ are independent and identically distributed with $E(\epsilon_{i,j}) = 0$ and $var(\epsilon_{i,j}) = 1$. As an exploratory exercise, the simple linear model of expression (2) was fit to data from each jurisdiction separately. Histograms of the resulting regression coefficients and their associated correlations are presented in Figure 8.

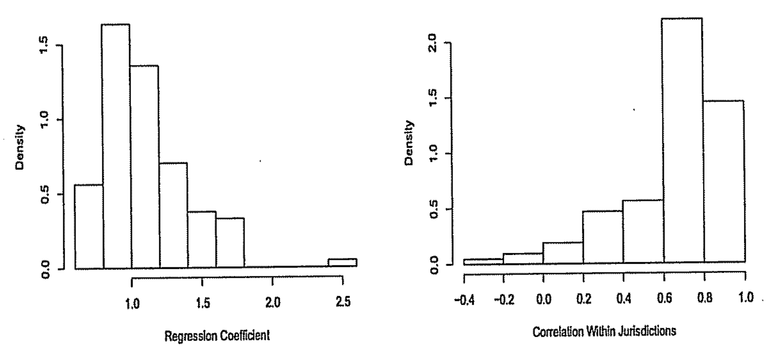


Figure 8: Histograms of estimated regression coefficients for model (2) fit to individual jurisdictions (left panel) and the associated sample correlations (right panel).

ANSWER QUESTION 6 NOW

An alternative model takes each jurisdiction to have a potentially unique regression. With $Y_{i,j}$ and $x_{i,j}$ defined as previously, take

$$Y_{i,j} = \beta_i x_{i,j} + \sigma_i \epsilon_{i,j}, \tag{3}$$

where now the error distribution is completely specified as $\epsilon_{i,j} \sim \text{iid}N(0,1)$. This model lacks an intercept term because a property with zero value should be assessed as such and this reduces the number of parameters that must be dealt with. We then use mixing distributions to model the $\{\beta_i : i = 1, \ldots, K\}$ and the $\{\sigma_i^2 : i = 1, \ldots, K\}$ as, for $\alpha > 0$, $\eta > 0$, $\lambda > 0$, and $\psi > 0$,

$$\begin{aligned} \beta_i &\sim \text{iid Gamma}(\alpha, \eta) \\ \sigma_i^2 &\sim \text{iid Inv Gamma}(\lambda, \psi) \end{aligned} \tag{4}$$

In particular, the form of density functions to be used are

$$\begin{aligned} g_1(\beta_i|\alpha,\eta) &= \frac{\eta^\alpha}{\Gamma(\alpha)} \beta_i^{\alpha-1} \exp(-\eta\beta_i); \quad \beta_i > 0 \\ &\text{and} \\ g_2(\sigma_i^2|\lambda,\psi) &= \frac{\psi^\lambda}{\Gamma(\lambda)} \frac{1}{(\sigma_i^2)^{\lambda+1}} \exp(-\psi/\sigma_i^2); \quad \sigma_i^2 > 0 \end{aligned} \tag{5}$$

ANSWER QUESTIONS 7 and 8 NOW

4 Questions

1. Based on the exploratory graphs and summary statistics given in Section 2, motivate the form of model (1). In particular,

   (a) What motivates the overall structure of model (1) as a mixture with random data model parameters λ_i; $i = 1, \ldots, K$?

   (b) What motivates the choice of a normal distribution for the data model?

   (c) What motivates the specification of data model variances as proportional to a power of data model means?

   (d) What motivates the choice of a lognormal distribution for the mixing distribution (or random parameter model)?

2. Relative to the choice of a normal distribution for the $R_{i,j}$ in model (1), what property of the actual data will we need to be mindful of to determine whether or not normal distributions will be appropriate for the problem? What parameter in the model will be central to ensuring that normal distributions can be used in this problem?

3. Suppose we would like to simulate data from model (1) to determine if it is capable of producing observations similar to the actual data. Determine numerical values that might be used for μ, σ^2, θ, and γ in the simulation.

   You may use without derivation the fact that, if Y has a lognormal distribution with parameters $-\infty < \mu < \infty$ and $\sigma^2 > 0$, then

$$
\begin{aligned}
E(Y) &= \exp(\mu + 0.5\sigma^2) \\
var(Y) &= \exp(2\mu + \sigma^2)\{\exp(\sigma^2) - 1\} \\
E(Y^k) &= \exp(k\mu + 0.5k^2\sigma^2)
\end{aligned}
$$

4. Find the marginal means, variances and covariances of the $R_{i,j}$ in model (1).

5. Beginning with the notation that $f_i(y_{i,j}|\lambda_i, \gamma)$ are the normal data model densities and $g(\lambda_i|\mu, \sigma^2)$ is the lognormal mixing distribution in model (1), write the marginal log likelihood that would be needed for maximum likelihood estimation in this problem.

6. Do the exploratory plots of Figures 5 through 8 support or fail to support the simple model of expression (2), or do they provide no information about this?

7. Suppose we will take a Bayesian approach to analysis of the model in expressions (3), (4), and (5).

   (a) What remaining quantities in the model will need to be assigned prior distributions?

   (b) Using $p(\theta|\cdot)$ to denote a full conditional posterior for the quantity θ, list the distributions we will need to simulate from if we use a Gibbs sampling algorithm to simulate values from the joint posterior. How many of these distributions will there be?

8. Suppose that all prior distributions not previously specified in the model will be taken as improper priors, that is, as uniform distributions on the parameter spaces.

Give explicit forms for each of the distributions you listed in part b) of question 7. Identify those which correspond to known distributional forms (e.g., normal or gamma) and simplify to the degree possible those that do not. Which will be easy to simulate from, and which will be more difficult?

These are a sketch of the answers hoped for. Other possibilities might exist for some of the questions that would be entirely adequate if they are both technically correct and logically consistent.

Question 1. (a) The overall mixture model is motivated in part by the problem, and in part by the exploratory plots presented in Figures 1, 2, and 3. In particular,

- There is a natural grouping mechanism because each jurisdiction has its own assessor and we would naturally suspect that the individual characteristics of these assessors will impact the relation between assessed values and sale prices.
- That different jurisdictions (or assessors) have individual characteristics that impact the problem is supported by Figures 1 and 2. The observed ratios in the individual jurisdictions shown in Figure 2 differ in at least location and variability, and none of the individual histograms in Figure 2 reflect the distributional form of the marginal histogram of Figure 1.
- The left side histogram of Figure 3, having one value for each jurisdiction, seems to reflect the same behavior as the marginal histogram of Figure 1 indicating that the overall shape of the marginal distribution may be produced primarily by variability among jurisdictions rather than within jurisdictions.

(b) The choice of a normal distribution for the data model is motivated primarily by the histograms of Figure 2 which are certainly more symmetric than the marginal histogram of Figure 1. The variances of these normal distributions will need to be small enough to avoid having the model place meaningful probability on the negative line, but, from the line plot of Figure 2 this would not seem to be an obvious problem.

(c) Specification of the data model variances as proportional to a power of data model means is motivated by the plots of Figure 4, particularly the Box-Cox plot on the right side of this figure.

(d) Choice of a lognormal mixing distribution is motivated by the shape of the histograms of Figure 1 and the left side of Figure 3.

Question 2. The property of the data that we will need to keep in mind is that all ratios must be positive. As noted in answer 1(d), the appropriateness of normal distributions for the data model will be determined by whether or not the variances are small enough to place no more than neglible probability on the negative line. While several parameters may be involved in this, a central role is played by the parameter γ in expression (1).

Question 3. The parameters needed might be chosen as follows.

(a) Values of θ and γ can be determined from the least squares fit to the Box-Cox plot of Figure 4, which had an intercept of about -2.8 and a slope of about 2.0. Let v_i denote a variance and m_i a mean. If log standard deviations are linear in log means, then,

$$0.5\log(v_i) = \beta_0 + \beta_1 m_i \Rightarrow v_i = exp(2\beta_0)\, m_i^{2\beta_1}$$

Thus, choose $\gamma = \exp(2\beta_0) = \exp(-5.6) = 0.0037$ and $\theta = 2\beta_1 = 4.0$.

(b) Values of μ and σ^2 can be selected by matching moments of the lognormal model with empirical moments of the jurisdiction mean ratios as presented in Table 1. Specifically, with s^2 and $\bar{x}$ denoting the mean and variance of the mean ratios across jurisdictions, we want

$$\begin{aligned} \exp(\mu + (1/2)\sigma^2) &= \bar{x} \\ \exp(2\mu + \sigma^2)\{\exp(\sigma^2) - 1\} &= s^2 \end{aligned}$$

Solving these equations with $\bar{x} = 1.1085$ and $s^2 = 0.0934$ from Table 1 results in $\mu = 0.0733$ and $\sigma^2 = 0.0664$.

Question 4. Marginal moments for this model are as follows,

$$\begin{aligned} E(R_{i,j}) &= E\{E(R_{i,j}|\lambda_i)\} = E(\lambda_i) = \exp(\mu + (1/2)\sigma^2) \\ var(R_{i,j}) &= var\{E(R_{i,j}|\lambda_i)\} + E\{var(R_{i,j}|\lambda_i)\} \\ &= var(\lambda_i) + E(\gamma\lambda_i^{\theta}) \\ &= \exp(2\mu + \sigma^2)\{\exp(\sigma^2) - 1\} + \gamma\left[\exp(\theta\mu + 0.5\theta^2\sigma^2)\right] \end{aligned}$$

Also,

$$E(R_{i,j}R_{i,k}) = E\{E(R_{i,j}R_{i,k}|\lambda_i)\} = E(\lambda_i^2) = \exp(2\mu + \sigma^2)\exp(\sigma^2)$$

which implies that

$$cov(Y_{i,j}, Y_{i,k}) = \exp(2\mu + \sigma^2)\{\exp(\sigma^2) - 1\}$$

and all other covariances are zero.

Question 5. First, let

$$f(\boldsymbol{y}_i|\lambda_i, \gamma) = \prod_{j=1}^{n_i} f(y_{i,j}|\lambda_i, \gamma)$$

and

$$h(\boldsymbol{y}_i|\gamma, \mu, \sigma^2) = \int_{-\infty}^{\infty} f(\boldsymbol{y}_i|\lambda_i, \gamma) g(\lambda_i|\mu, \sigma^2)\, d\lambda_i$$

Then the log likelihood needed is

$$\ell(\mu, \sigma^2, \gamma) = \sum_{i=1}^{K} \log\{h(\boldsymbol{y}_i|\gamma, \mu, \sigma^2)\}$$

The key element of this is that we need to work with the joint distributions of the $Y_{i,j}$ within each jurisdiction because these variables are not independent (for a common index i).

Question 6. While Figure 5 would suggest that this model might be fit to the jurisdication means, Figures 7 and 8 suggest that data within jurisdictions may have relations between assessed values and sale prices that do not all follow this same model. In particular, the left hand of Figure 8 indicates that regressions of assessed values on sale prices vary considerably across jurisdictions, and the right hand side of Figure 8 suggests that the strength of (linear) relation also varies.

Question 7. (a) We will need to assign prior distributions to the parameters α, η, λ and ψ.

(b) Here, it will be most convenient to include the β_i and σ_i^2 in the simulation so that the full conditional distributions we will need to simulate from are $p(\alpha|\cdot)$, $p(\eta|\cdot)$, $p(\lambda|\cdot)$, $p(\psi|\cdot)$, $\{p(\beta_i|\cdot) : i = 1, \dots, K\}$ and $\{p(\sigma_i^2|\cdot) : i = 1, \dots, K\}$ for a total of $2K + 4$ distributions.

Question 8. Full conditional posteriors are derived as follows.

(a) $p(\alpha|\cdot)$.

$$\begin{aligned} p(\alpha|\cdot) &\propto \pi(\alpha) \prod_{i=1}^{K} g_1(\beta_i|\alpha, \eta) \\ &\propto \left\{ \frac{\eta^\alpha}{\Gamma(\alpha)} \right\}^K \prod_{i=1}^{K} \beta_i^{\alpha-1}. \end{aligned}$$

(b) $p(\eta|\cdot)$.

$$\begin{aligned} p(\eta|\cdot) &\propto \pi(\eta) \prod_{i=1}^{K} g_1(\beta_i|\alpha, \eta) \\ &\propto \eta^{K\alpha} \exp\{(-\textstyle\sum \beta_i)\eta\}, \end{aligned}$$

which is a gamma distribuiton with parameters $K\alpha + 1$ and $\sum \beta_i + 1$.

(c) $p(\lambda|\cdot)$.

$$\begin{aligned} p(\lambda|\cdot) &\propto \pi(\lambda) \prod_{i=1}^{K} g_2(\sigma_i^2|\lambda, \psi) \\ &\propto \left\{ \frac{\psi^\lambda}{\Gamma(\lambda)} \right\}^K \prod_{i=1}^{K} \frac{1}{(\sigma_i^2)^{\lambda+1}}. \end{aligned}$$

(d) $p(\psi|\cdot)$.

$$\begin{aligned} p(\psi|\cdot) &\propto \pi(\psi) \prod_{i=1}^{K} g_2(\sigma_i^2|\lambda, \psi) \\ &\propto \psi^{K\lambda} \exp\{-\psi/(\textstyle\sum \sigma_i^2)\}, \end{aligned}$$

which is a gamma distribution with parameters $K\lambda + 1$ and $1 + 1/\sum \sigma_i^2$.

(e) $p(\sigma_i^2|\cdot)$ for $i = 1, \ldots, K$.

$$\begin{aligned} p(\sigma_i^2|\cdot) &\propto g_2(\sigma_i^2|\lambda,\psi) f(\boldsymbol{y}_i|\beta_i,\sigma_i^2) \\ &\propto \frac{1}{(\sigma_i^2)^{\lambda+1}} \exp(-\psi/\sigma_i^2) \frac{1}{(\sigma_i^2)^{n_i/2}} \exp\left\{-\frac{1}{2\sigma_i^2}\sum_{j=1}^{n_i}(y_{i,j}-\beta_i x_{i,j})^2\right\} \\ &= \frac{1}{(\sigma_i^2)^{\lambda+n_i/2+1}} \exp\left\{-\frac{\psi}{2}\sum(y_{i,j}-\beta_i x_{i,j})^2 \,/\, \sigma_i^2\right\}, \end{aligned}$$

which is an inverse gamma distribution with parameters $\lambda+n_i/2$ and $(\psi/2)\sum(y_{i,j}-\beta_i x_{i,j})^2$.

(f) $p(\beta_i|\cdot)$ for $i = 1, \ldots, K$.

$$\begin{aligned} p(\beta_i|\cdot) &\propto g_1(\beta_i|\alpha,\eta) f(\boldsymbol{y}_i|\beta_i,\sigma_i^2) \\ &\propto \beta_i^{\alpha-1}\exp(-\eta\beta_i)\exp\left\{-\frac{1}{2\sigma_i^2}\sum_{j=1}^{n_i}(y_{i,j}-\beta_i x_{i,j})^2\right\} \end{aligned}$$

The distributons for η, ψ and $\{\sigma_i^2 : i = 1, \ldots, K\}$ will be easy to simulate from while those for α, λ and $\{\beta_i : i = 1, \ldots, K\}$ will be more difficult.