CASUALTY ACTUARIAL SOCIETY  
AND THE  
CANADIAN INSTITUTE OF ACTUARIES

MAS-II  
Modern Actuarial Statistics II

October 25, 2019

INSTRUCTIONS TO CANDIDATES

1. This 84 point examination consists of 42 multiple choice questions each worth 2 points.

2. To answer the multiple choice questions, use the short-answer card provided and a number 2 or HB pencil only.
   
   - Fill in that it is Fall 2019 and that the exam name is MAS-II.
   
   - Darken the spaces corresponding to your Candidate ID number. Five rows are available. If your Candidate ID number is fewer than 5 digits, include leading zeros. For example, if your Candidate ID number is 987, consider that your Candidate ID number is 00987, enter a zero on the first row, a zero on the second row, 9 on the third row, 8 on the fourth row, and 7 on the fifth [last] row. Write in your Candidate ID number next to the place where you darken the spaces for your Candidate ID number. Your name, or any other identifying mark, must not appear on the short-answer card.

   - Mark your short-answer card during the examination period. No additional time will be allowed for this after the exam has ended. Make your marks dark and fill in the spaces completely.

   - For each of the multiple choice questions, select the one best answer and fill in the corresponding letter. One quarter of the point value of the question will be subtracted for each incorrect answer. No points will be added or subtracted for responses left blank.

3. Do all problems until you reach the last page of the examination where "END OF EXAMINATION" is marked.

CONTINUE TO NEXT PAGE OF INSTRUCTIONS
©2019 Casualty Actuarial Society
4. Prior to the start of the exam you will have a **fifteen-minute reading period** in which you can silently read the questions and check the exam booklet for missing or defective pages. Writing will NOT be permitted during this time and you will not be permitted to hold pens or pencils. You will also not be allowed to use calculators. The supervisor has additional exams for those candidates who have defective exam booklets.

- Verify that you have a copy of the case study, “Warranty Payments Case Study”, included in your exam packet.
- Verify that you have a copy of “Tables for CAS MAS-II” included in your exam packet.
- Candidates should not interpolate in the tables unless explicitly instructed to do so in the problem, rather, a candidate should round the result that would be used to enter a given table to the same level of precision as shown in the table and use the result in the table that is nearest to that indicated by rounded result. For example, if a candidate is using the Tables of the Normal Distribution to find a significance level and has a Z value of 1.903, the candidate should round to 1.90 to find cumulative probability in the Normal table.

5. Your Examination Envelope is pre-labeled with your Candidate ID number, name, exam number, and test center. **Do not remove this label.** Keep a record of your Candidate ID number for future inquiries regarding this exam.

6. Candidates must remain in the examination center until two hours after the start of the examination. The examination starts after the reading period is complete. You may leave the examination room to use the restroom with permission from the supervisor. To avoid excessive noise during the end of the examination, candidates may not leave the exam room during the last fifteen minutes of the examination.

7. **At the end of the examination, place the short-answer card in the Examination Envelope.** Nothing written in the examination booklet will be graded. **Only the short-answer card will be graded.** Also place any included reference materials in the Examination Envelope. **BEFORE YOU TURN THE EXAMINATION ENVELOPE IN TO THE SUPERVISOR, BE SURE TO SIGN IT IN THE SPACE PROVIDED ABOVE THE CUT-OUT WINDOW.**

8. If you have brought a self-addressed, stamped envelope, you may put the examination booklet and scrap paper inside and submit it separately to the supervisor. It will be mailed to you. **Do not put the self-addressed stamped envelope inside the Examination Envelope.** Interoffice mail is not acceptable.

   If you do not have a self-addressed, stamped envelope, please place the examination booklet in the Examination Envelope and seal the envelope. You may not take it with you. **Do not put scrap paper in the Examination Envelope.** The supervisor will collect your scrap paper.

Candidates may obtain a copy of the examination from the CAS Web Site.

All extra answer sheets, scrap paper, etc. must be returned to the supervisor for disposal.

**CONTINUE TO NEXT PAGE OF INSTRUCTIONS**
9. Candidates must not give or receive assistance of any kind during the examination. Any cheating, any attempt to cheat, assisting others to cheat, or participating therein, or other improper conduct will result in the Casualty Actuarial Society and the Canadian Institute of Actuaries disqualifying the candidate's paper, and such other disciplinary action as may be deemed appropriate within the guidelines of the CAS Policy on Examination Discipline.

10. The exam survey is available on the CAS Web Site in the “Admissions/Exams” section. Please submit your survey by November 8, 2019.

END OF INSTRUCTIONS
1.

In the supplemental material, you have been given a case study, “Warranty Payments Case Study”. You have also been asked to evaluate three statements regarding the results of the study.

   I. Model I is a marginal model.
   II. Placing the Store variable in the random effects category is appropriate since we only have a subset of possible results.
   III. Model III is a nested model of Model IV.

Determine which of the above statements are true.

A. I only
B. II only
C. III only
D. I, II and III
E. The answer is not given by (A), (B), (C), or (D)
In the supplemental material, you have been given a case study, “Warranty Payments Case Study”. You have also been asked to evaluate three statements regarding the residuals results of the study.

I. Model II does not show any indication of heteroscedasticity.
II. Model III does show indications of heteroscedasticity.
III. Model IV does show indications of heteroscedasticity.

Determine which of the above statements are true.

A. I only
B. II only
C. III only
D. I, II and III
E. The answer is not given by (A), (B), (C), or (D)
An actuary is constructing a credibility formula to apply to claim severity. To visualize the partial credibility, a plot showing credibility as a function of claim counts is created.

Determine which credibility method the actuary is using.

A. Limited-fluctuation credibility with a full-credibility standard of 1300 claims.
B. Limited-fluctuation credibility with a full-credibility standard of 800 claims.
C. Limited-fluctuation credibility with $\lambda_F = 541$ and $C_X = 1.053$.
D. Bühlmann credibility with EVPV = 0.266 and VHM = 0.016.
E. Bühlmann credibility with EVPV = 0.844 and VHM = 0.008.
4.

You are given the following parameters.

- Assume the full-credibility standard using limited-fluctuation credibility is based on $\alpha = 0.05$ and $k = 0.02$.
- The expected claim frequency per exposure unit is 0.03.
- $W$ is the full-credibility standard for claim frequency in exposure units assuming Poisson claim frequency.
- $V$ is the full-credibility standard for claim frequency in exposure units assuming binomial claim frequency.

Calculate $|W - V|$.

A. Less than 4,000
B. At least 4,000 but less than 6,000
C. At least 6,000 but less than 8,000
D. At least 8,000 but less than 10,000
E. At least 10,000
An insurance company writes insurance in three states. The company experiences the following aggregate loss distributions over the past eight years:

<table>
<thead>
<tr>
<th>State</th>
<th>Mean Annual Loss (in millions)</th>
<th>Standard Deviation of Annual Loss (in millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>147</td>
<td>72</td>
</tr>
<tr>
<td>V</td>
<td>57</td>
<td>27</td>
</tr>
<tr>
<td>W</td>
<td>54</td>
<td>24</td>
</tr>
</tbody>
</table>

Assume that the volume of business and the true underlying size of the average loss have not changed for these eight years and will continue to remain stable next year.

Calculate the estimate of next year’s aggregate loss for State W using the nonparametric empirical Bayes method.

A. Less than 60 million
B. At least 60 million, but less than 65 million
C. At least 65 million, but less than 70 million
D. At least 70 million, but less than 75 million
E. At least 75 million
6.

Policies belong to one of two possible risk groups, Risk Group R and Risk Group S. You are given the following information:

- 40% of policies belong to Risk Group R.
- Risk Group R has claim frequencies that are Poisson distributed with \( \lambda = 3 \).
- Risk Group R has claim severity that is uniformly distributed between 3000 and 5000.

- 60% of policies belong to Risk Group S.
- Risk Group S has claim frequencies that are Poisson distributed with \( \lambda = 1 \).
- Risk Group S has claim severity that is uniformly distributed between 50 and 500.

- Claim frequency and claim severity are independently distributed given a risk group.

The Bühlmann credibility method is used to calculate the next year’s predicted aggregate loss given two prior years of loss experience for a given risk.

Calculate the Bühlmann credibility factor for this risk.

A. Less than 0.70
B. At least 0.70, but less than 0.75
C. At least 0.75, but less than 0.80
D. At least 0.80, but less than 0.85
E. At least 0.85
The heights of 20 schoolgirls are measured on a yearly basis from age 6 to 10. The girls were classified into three groups. The three groups are based on the height of their mother (SHORT, MEDIUM, or TALL).

The table below illustrates the data:

<table>
<thead>
<tr>
<th>Child ID</th>
<th>Height of Mother</th>
<th>Child Age</th>
<th>Child Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SHORT</td>
<td>6</td>
<td>111</td>
</tr>
<tr>
<td>1</td>
<td>SHORT</td>
<td>7</td>
<td>116.4</td>
</tr>
<tr>
<td>1</td>
<td>SHORT</td>
<td>8</td>
<td>121.7</td>
</tr>
<tr>
<td>1</td>
<td>SHORT</td>
<td>9</td>
<td>126.3</td>
</tr>
<tr>
<td>1</td>
<td>SHORT</td>
<td>10</td>
<td>130.5</td>
</tr>
<tr>
<td>1</td>
<td>SHORT</td>
<td>6</td>
<td>110</td>
</tr>
<tr>
<td>......</td>
<td>......</td>
<td>......</td>
<td>......</td>
</tr>
<tr>
<td>7</td>
<td>MEDIUM</td>
<td>6</td>
<td>116</td>
</tr>
<tr>
<td>......</td>
<td>......</td>
<td>......</td>
<td>......</td>
</tr>
<tr>
<td>20</td>
<td>TALL</td>
<td>10</td>
<td>152.3</td>
</tr>
</tbody>
</table>

The hierarchical Linear Mixed Model given below is used to model $HEIGHT_{ti}$, where $HEIGHT_{ti}$ is the $t^{th}$ annual height measurement of the $i^{th}$ schoolgirl.

Level 1 Model (Time):

$$HEIGHT_{ti} = b_{0i} + b_{1i} \times AGE_{ti} + \varepsilon_{ti}$$

where $\varepsilon_{ti} \sim N(0, \sigma^2)$

Level 2 Model (individual schoolgirl):

$$b_{0i} = \beta_0 + \beta_2 \times SHORT_i + \beta_3 \times TALL_i + u_{0i}$$

$$b_{1i} = \beta_1 + u_{1i}$$

where $u_t = \begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N(0, \mathbf{D})$

Determine which of the following Linear Mixed Models is equivalent to the above hierarchical model.

A. $HEIGHT_{ti} = \beta_0 + \beta_1 \times AGE_{ti} + \beta_2 \times SHORT_i + \beta_3 \times TALL_i + u_{0i} + \varepsilon_{ti}$

B. $HEIGHT_{ti} = \beta_0 + \beta_1 \times AGE_{ti} + \beta_2 \times SHORT_i + \beta_3 \times TALL_i + u_{0i} + u_{1i} + \varepsilon_{ti}$

C. $HEIGHT_{ti} = \beta_0 + \beta_1 \times AGE_{ti} + \beta_2 \times SHORT_i + \beta_3 \times TALL_i + u_{0i} + u_{1i} \times AGE_{ti} + \varepsilon_{ti}$

D. $HEIGHT_{ti} = \beta_0 + \beta_1 \times SHORT_i + \beta_2 \times TALL_i + \beta_3 \times AGE_{ti} + u_{0i} + u_{1i} \times AGE_{ti} + \varepsilon_{ti}$

E. $HEIGHT_{ti} = \beta_0 + \beta_1 \times AGE_{ti} + \beta_2 \times SHORT_i + \beta_3 \times TALL_i + u_{0i} \times AGE_{ti} + u_{1i} + \varepsilon_{ti}$

CONTINUED ON NEXT PAGE
You are given the following statements are true regarding the variance-covariance $D$ matrix relating the random effects in a Linear Mixed Model:

I. The matrix is always symmetrical.
II. All elements of the matrix must be non-negative.
III. Under the compound symmetry structure, all the elements along the diagonal of the matrix are equal.

Determine which of the above statements are true.

A. None of I, II, or III is true
B. I and II only
C. I and III only
D. II and III only
E. The answer is not given by (A), (B), (C), or (D)
9.

You are testing the relative efficiency of two types of gasoline in miles per gallon.

- You use six similar vehicles for your test.
- You perform the test on four different days.
- Each automobile only receives one of the two different types of gasoline.
- There is a random intercept associated with each automobile $i$.
- There is a random effect associated with each testing day $t$.

Determine the best structure for the Linear Mixed Model that describes the relationship between the covariates and miles per gallon.

A. $MPG_{ti} = \beta_0 + \beta_1 \times FUEL_{1i} + \epsilon_{ti}$

B. $MPG_{ti} = \beta_0 + \beta_1 \times FUEL_{1i} + u_{0i} + u_{1i} + \epsilon_{ti}$

C. $MPG_{ti} = \beta_0 + \beta_1 \times FUEL_{1i} + u_{0i} + u_{t0} + \epsilon_{ti}$

D. $MPG_{ti} = \beta_0 + \beta_1 \times FUEL_{1i} + \beta_2 \times TESTDAY_t + u_{0i} + \epsilon_{ti}$

E. $MPG_{ti} = \beta_0 + \beta_1 \times FUEL_{1i} + \beta_2 \times TESTDAY_t + u_{t0} + \epsilon_{ti}$
10.

Your company conducts a test to determine which software program is more efficient for new claims adjustors to use.

- There are two software programs tested.
- Twenty claims adjustors participated in this study; each adjustor is given a software program to use at random.
- Each claims adjustor uses the same software program for all claims the adjustor processed for an entire week.

You build a Linear Mixed Model to estimate the amount of time it takes a claims adjustor to handle a particular claim. You use the following variables in the model:

- Claims adjustor ID
- Software program used
- Tenure of the claims adjustor
- Type of claim

Determine which of the model variables is a random factor.

A. Claims adjustor ID
B. Software program used
C. Tenure of the claims adjustor
D. Type of claim
E. None of these variables are random factors
You are building a Linear Mixed Model to estimate the number of polices written by sales representatives in various agencies in the company. You are given the following information:

- The number of polices written \((Y_{ij})\) is measured for a sales representative \(i\) who works in agency \(j\).
- \(L_j\) and \(M_j\) represent characteristics of individual agencies.
- \(N_{ij}\) represents a characteristic of an individual sales representative.
- \(\mu_j\) is a random effect for each agency that is normally distributed with mean 0 and variance \(\sigma_{\text{agency}}^2\).
- \(\varepsilon_{ij}\) is a residual that is normally distributed with mean 0 and variance \(\sigma^2\).
- The general model specification is in the following form:

\[
Y_{ij} = \beta_0 + \beta_1 \times L_j + \beta_2 \times M_j + \beta_3 \times N_{ij} + \beta_4 \times M_j \times N_{ij} + u_j + \varepsilon_{ij}
\]

Determine the hierarchical model specification.

A. Level 1: \(Y_{ij} = b_{0j} + b_{1j} \times M_j + \varepsilon_{ij}\)
   Level 2: \(b_{0j} = \beta_0 + \beta_1 \times L_j + \beta_2 \times N_{ij} + u_j\)
   \(b_{1j} = \beta_3 + \beta_4 \times N_{ij}\)

B. Level 1: \(Y_{ij} = b_{0j} + b_{1j} \times M_j\)
   Level 2: \(b_{0j} = \beta_0 + \beta_1 \times L_j + \beta_2 \times N_{ij} + u_j + \varepsilon_{ij}\)
   \(b_{1j} = \beta_3 + \beta_4 \times N_{ij}\)

C. Level 1: \(Y_{ij} = b_{0j} + b_{1j} \times N_{ij} + \varepsilon_{ij}\)
   Level 2: \(b_{0j} = \beta_0 + \beta_1 \times L_j + \beta_2 \times M_j + u_j\)
   \(b_{1j} = \beta_3 + \beta_4 \times M_j\)

D. Level 1: \(Y_{ij} = b_{0j} + b_{1j} \times N_{ij}\)
   Level 2: \(b_{0j} = \beta_0 + \beta_1 \times L_j + \beta_2 \times M_j + u_j + \varepsilon_{ij}\)
   \(b_{1j} = \beta_3 + \beta_4 \times M_j\)

E. Level 1: \(Y_{ij} = b_{0j} + b_{1j} \times N_{ij} + \varepsilon_{ij}\)
   Level 2: \(b_{0j} = \beta_0 + \beta_1 \times L_j + \beta_2 \times M_j + u_j\)
   \(b_{1j} = \beta_2 + \beta_4 \times M_j\)

CONTINUED ON NEXT PAGE
12.

You are performing $F$-tests to test linear hypotheses about multiple fixed effects in a Linear Mixed Model. You use the Kenward-Roger method to approximate the denominator degrees of freedom. You are given the following statements about the Kenward-Roger adjustment:

I. This adjustment is necessary because the variance of the fixed effects is biased downward due to the estimation of the variance-covariance matrix.
II. This adjustment accounts for the presence of random effects and correlated residuals in the Linear Mixed Model.
III. In general, the adjustment is more critical for data sets with larger sample sizes than for smaller sample sizes.

Determine which of the proceeding statements are true.

A. I only
B. II only
C. III only
D. I, II and III
E. The answer is not given by (A), (B), (C), or (D)
13.

You are reviewing a dataset of automobile accidents using a Linear Mixed Model.
- Your null hypothesis is that the variance of the residuals remains constant across all four quarters (Quarter 1, Quarter 2, Quarter 3, and Quarter 4).
- The alternative hypothesis is that the variance of the residuals varies by quarter.
- To test your hypotheses, you use the log likelihood test using the chi-square distribution.

Determine the degrees of freedom used for the test.

A. 0.5 (average of 0 and 1)
B. 1.0
C. 2.0
D. 3.0
E. 4.0
Exam MAS-II Fall 2019

14.

You are conducting an experiment with two levels: packets of pea plant seeds and the pea plant seeds themselves. You are interested in the effect of different types of light on height of the pea plant after four weeks. Below is the final model:

\[
\text{HEIGHT}_{ij} = \beta_0 + \beta_1 \cdot \text{TREAT1}_j + \beta_2 \cdot \text{TREAT2}_j + \beta_3 \cdot \text{TREAT3}_j + \beta_4 \cdot \text{MASS}_{ij} \\
+ \beta_5 \cdot \text{NUMSEEDS}_j + u_j + \varepsilon_{ij}
\]

- TREAT1, TREAT2, and TREAT3 correspond to 3 of the 4 light types used
- MASS is the mass of the given seed in grams
- NUMSEEDS is the number of seeds in the seed packet
- The model output is given below:

```
###Fixed effects: height ~ treatment + mass + numseeds
#
###                         Value  Std.Error  DF
###(Intercept)          10.118    0.7067   294
###TREAT1               -5.320    0.3493    22
###TREAT2               -7.824    0.3388    22
###TREAT3               -4.373    0.3474    22
###mass                0.1110     0.1049   294
###numseeds            -0.053     0.0422    22
```

Determine the equation to predict pea plant height for a specific seed in a specific seed packet under TREAT1.

A. \( \text{HEIGHT}_{ij} = 4.798 + 0.111 \cdot \text{MASS}_{ij} - 0.053 \cdot \text{NUMSEEDS}_j + \tilde{u}_j + \varepsilon_{ij} \)
B. \( \text{HEIGHT}_{ij} = 4.798 + 0.111 \cdot \text{MASS}_{ij} - 0.053 \cdot \text{NUMSEEDS}_j + \tilde{u}_j \)
C. \( \text{HEIGHT}_{ij} = 4.798 + 0.111 \cdot \text{MASS}_{ij} - 0.053 \cdot \text{NUMSEEDS}_j \)
D. \( \text{HEIGHT}_{ij} = 10.118 + 0.111 \cdot \text{MASS}_{ij} - 0.053 \cdot \text{NUMSEEDS}_j + \tilde{u}_j + \varepsilon_{ij} \)
E. \( \text{HEIGHT}_{ij} = 10.118 + 0.111 \cdot \text{MASS}_{ij} - 0.053 \cdot \text{NUMSEEDS}_j + \tilde{u}_j \)
You are given the following statements about a Linear Mixed Model and Marginal Model implied by that Linear Mixed Model:

I. When scoring the observations in the modeling dataset, both models will yield the same conditional predicted values.
II. Even if the variance-covariance matrix for the marginal residual errors in the Implied Marginal Model is nonpositive-definite, you may still be able to fit the Linear Mixed Model directly.
III. The interpretation of the covariance parameters is the same between both models.

Determine which of the proceeding statements are true.

A. None of I, II, or III is true
B. I and II only
C. I and III only
D. II and III only
E. The answer is not given by (A), (B), (C), or (D)
Quadratic approximation is employed to estimate the posterior distribution for a model with two correlated parameters, $\theta_1$ and $\theta_2$. Summary statistics of marginal posterior distributions are provided below.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>9.04</td>
<td>10.24</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>2.56</td>
<td>0.82</td>
</tr>
<tr>
<td>$\theta_1 - \theta_2$</td>
<td>6.30</td>
<td>11.12</td>
</tr>
</tbody>
</table>

Let $x$ be the lower bound of the 96% HPDI for $\theta_1$ and $y$ be the upper bound of the 96% HPDI for $\theta_2$.

Calculate $x - y$.

A. Less than -3.0
B. Greater than or equal to -3.0 but less than -1.0
C. Greater than or equal to -1.0 but less than 1.0
D. Greater than or equal to 1.0 but less than 3.0
E. Greater than or equal to 3.0
A team of data scientists is analyzing renewal rates on personal auto policies for insurance company XYZ. Using historical renewal data and a binomial distribution, they have calculated the posterior distribution displayed below:

![Density vs Renewal Rate Graph](image)

Identify which of the following ranges is most likely the 30% highest posterior density interval for renewal rate.

A. [0.00, 0.30]
B. [0.15, 0.85]
C. [0.40, 0.80]
D. [0.60, 0.70]
E. [0.70, 1.00]
You wish to model the effect of gender on a random count outcome variable \(X\). To achieve this, the variable for gender, denoted by \(F\), is transformed such that it takes the value of 1 for females, otherwise it takes the value of 0. You fit the following model to your data:

\[
X_i \sim \text{Poisson}(\lambda_i) \\
\log(\lambda_i) = \alpha + \beta_F F_i \\
\alpha \sim \text{Normal}(0, 10) \\
\beta_F \sim \text{Normal}(0, 10)
\]

You are given the following output from the model:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>2</td>
<td>0.5</td>
<td>1.9</td>
</tr>
<tr>
<td>( \beta_F )</td>
<td>0.5</td>
<td>0.1</td>
<td>0.5</td>
</tr>
</tbody>
</table>

You obtain the following samples from the posterior distribution:

<table>
<thead>
<tr>
<th>Sample</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>1.9</td>
<td>1.8</td>
<td>2.1</td>
<td>2.1</td>
<td>2.2</td>
<td>2.1</td>
<td>1.7</td>
</tr>
<tr>
<td>( \beta_F )</td>
<td>0.4</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.7</td>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Calculate the estimate for the posterior median of the average female count based on this information.

A. Less than 2.5  
B. Greater than or equal to 2.5 but less than 7.5  
C. Greater than or equal to 7.5 but less than 12  
D. Greater than or equal to 12 but less than 13  
E. Greater than or equal to 13
You are considering attempting a college entrance exam and would like to understand the percentage of candidates that successfully pass the exam. The alumni have provided you with varying estimates of the percentage of successful candidates. You summarize their responses in the table below:

<table>
<thead>
<tr>
<th>Alumni ID</th>
<th>Years since passing the entrance exam</th>
<th>Pass Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>70%</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>60%</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>50%</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>40%</td>
</tr>
</tbody>
</table>

You weight their responses using the reciprocal of the number of years since they passed their entrance exam and this forms your prior distribution of the exam pass rate.

You then poll three graduating seniors from your school and find out that two of those students were successful in passing the exam.

You want to use grid approximation to describe the posterior distribution.

Calculate the posterior probability that the pass rate is 70%.

A. Less than 50%
B. Greater than or equal to 50% but less than 55%
C. Greater than or equal to 55% but less than 60%
D. Greater than or equal to 60% but less than 65%
E. Greater than or equal to 65%
Three hierarchical Bayesian regression models, Model I, Model II, and Model III, are fit to the same dataset. The dataset contains information that relates whether or not an insurance policy sale was completed and includes four potential covariates:

1. region, $i$, of the sales office where $i \in \{1, 2, \ldots, 7\}$
2. The type of product, $j$, sold where $j \in \{1, 2, \ldots, 6\}$
3. The experience level of the insurance sales representative
4. Whether the sale attempt was made as part of a bundle of other products or made in isolation

All three models’ fits include experience and bundle as covariates. Model II also includes region and Model III also includes both region and product.

Model III is specified as follows:

$$
    y_i \sim \text{Bernoulli}(p_i) \\
    \text{logit}(p_i) = b + b_{\text{region}[i]} + b_{\text{product}[j]} + (b_1 + b_2 \times x_{\text{bundle}}) \times x_{\text{experience}} \\
    b_{\text{region}} \sim \text{Normal}(0, \sigma_{\text{region}}) \\
    b_{\text{product}} \sim \text{Normal}(0, \sigma_{\text{product}}) \\
    b \sim \text{Normal}(0, 10) \\
    b_1 \sim \text{Normal}(0, 10) \\
    b_2 \sim \text{Normal}(0, 10) \\
    \sigma_{\text{region}} \sim \text{HalfCauchy}(0, 1) \\
    \sigma_{\text{product}} \sim \text{HalfCauchy}(0, 1)
$$

You are given the following comparison between Model I and Model II.

<table>
<thead>
<tr>
<th>Model</th>
<th>WAIC</th>
<th>pWAIC</th>
<th>dWAIC</th>
<th>Weight</th>
<th>SE</th>
<th>dSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model I</td>
<td>682.4</td>
<td>2.9</td>
<td>151.2</td>
<td>0.0</td>
<td>9.45</td>
<td>18.59</td>
</tr>
<tr>
<td>Model II</td>
<td>531.2</td>
<td>8.1</td>
<td>0.0</td>
<td>1.0</td>
<td>19.43</td>
<td>NA</td>
</tr>
</tbody>
</table>

You are also provided the posterior means and 90% Bayesian credible intervals for the parameters in Model III in the graphic on the following page.

Question #20 continued on the next page.
Determine the most likely value of pWAIC for Model III.

A. 5.0
B. 8.0
C. 11.0
D. 14.0
E. 17.0
A regression model for the number of automobile accidents, $Y_i$, for driver $i$ assumes a zero-inflated Poisson (ZIPoisson) distribution with $p_i$ being the probability of zero. The model is shown below.

\[
\begin{align*}
Y_i & \sim \text{ZIPoisson}(\lambda_i, p_i) \\
\ln(\lambda_i) & = \beta_0 + \beta_1 x_{1i} \\
\text{logit}(p_i) & = \alpha_0 + \alpha_1 x_{1i} \\
\beta_0 & \sim \text{Normal}(0, 10) \\
\beta_1 & \sim \text{Normal}(0, 1) \\
\alpha_0 & \sim \text{Normal}(0, 2) \\
\alpha_1 & \sim \text{Normal}(0.5, 0.5)
\end{align*}
\]

After sufficient warmup, two samples (sample 1 and sample 2) from the posterior distribution are taken and provided in the table below:

<table>
<thead>
<tr>
<th></th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample 1</td>
<td>-0.96</td>
<td>0.13</td>
<td>-2.75</td>
<td>0.90</td>
</tr>
<tr>
<td>sample 2</td>
<td>0.55</td>
<td>-0.23</td>
<td>-0.63</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Compute the mean probability of a driver having 0 automobile accidents when $x_1 = 2.0$ from the posterior predictive distribution.

A. Less than 0.75  
B. Greater than or equal to 0.75 but less than 0.80  
C. Greater than or equal to 0.80 but less than 0.85  
D. Greater than or equal to 0.85 but less than 0.90  
E. Greater than or equal to 0.90

CONTINUED ON NEXT PAGE
22.

Your coworker built a linear model predicting a random variable $Y$:

$$Y_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta_{x_1} \log(x_{1i}) + \beta_{x_2} x_2$$

$$\alpha \sim \text{Normal}(100, 100)$$

$$\beta_{x_1} \sim \text{Normal}(0, 5)$$

$$\beta_{x_2} \sim \text{Normal}(0, 5)$$

$$\sigma \sim \text{Uniform}(0, 100)$$

You are given additional detail about pre-processing performed on the predictor variables:

- All variables are continuous numeric
- $x_{1i}$ is strictly positive
- Variable $x_{2i}$ has been standardized to have mean 0 and standard deviation 1

Your coworker gives you the following three statements:

I. $\alpha =$ average outcome
II. $\beta_{x_1} =$ change in outcome for each 1 log change of $x_{1i}$
III. $\beta_{x_2} =$ change in outcome for each 1 standard deviation change of $x_{2i}$

Determine which of the above statements are false.

A. I only
B. II only
C. III only
D. I, II and III
E. The answer is not given by (A), (B), (C), or (D)
You are given the following statements to evaluate regarding the benefits of using a multilevel approach in your Bayesian model:

I. They provide improved estimates when different clusters have a different number of observations in the training dataset.
II. Estimating from the posterior distribution is simpler.
III. The models produced are always easier to understand.

Determine which of the above statements are benefits of a Bayesian multilevel modelling approach.

A. I only
B. II only
C. III only
D. I, II and III
E. The answer is not given by (A), (B), (C), or (D)
24.

Given the following types of Bayesian model diagnostic tools:

I. Leave one out cross validation
II. WAIC
III. Counterfactual plots

Determine which of the above diagnostics tools can rely on data not included in the training data set.

A. I only
B. II only
C. III only
D. I, II and III
E. The answer is not given by (A), (B), (C), or (D)
25.

A book of 122 commercial policies are observed for one year. The observed claim counts distribution is shown below:

<table>
<thead>
<tr>
<th>Claim count</th>
<th>Policies</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

Calculate the log-cumulative-odds of 3 claims.

A. Less than -2.0  
B. Greater than or equal to -2.0 but less than 0.0  
C. Greater than or equal to 0.0 but less than 2.0  
D. Greater than or equal to 2.0 but less than 4.0  
E. Greater than or equal to 4.0
You observe a linear relationship between the amount of snowfall in a territory and the number of collision claims, $Y$, filed by insured drivers in that territory. You describe that relationship as follows:

$$ Y \sim \text{Normal}(\mu, \sigma) $$

$$ \mu = \beta_0 + \beta_1 \cdot \text{snowfall} $$

You decide to model this relationship using linear regression with Ordinary Least Squares (OLS) and compare it to the same model fit using Bayesian Markov chain Monte Carlo (MCMC), with appropriate priors.

You are given the following statements regarding your models:

I. The MCMC model is specified with one likelihood function and three prior distributions.
II. Parameter estimates from an OLS model and MCMC model will yield similar estimates when using flat priors.
III. The OLS regression provides information on the uncertainty of the estimate of the sigma parameter ($\sigma$).

Determine which of the above statements are true about your models.

A. None of I, II, or II is true
B. I and II only
C. I and III only
D. II and III only
E. The answer is not given by (A), (B), (C), or (D)
27.

When modeling claim frequency, an actuary chooses to use the following Bayesian model:

\[ Y \sim \text{Poisson}(\mu) \]
\[ \mu \sim \text{Gamma}(\alpha = 1, \theta = 0.1) \]

Suppose the actuary observed 10 claims from 250 policies in the past year.

Calculate the mean of the posterior predictive distribution for 250 policies.

A. Less than 5
B. Greater than or equal to 5 but less than 10
C. Greater than or equal to 10 but less than 15
D. Greater than or equal to 15 but less than 20
E. Greater than or equal to 20
A varying-intercepts model is built to model the test scores of students in eight schools:

\[ \text{score}_i \sim \text{Normal}(\mu_i, p_i) \]
\[ \mu_i = \alpha_{\text{SCHOOL}[i]} + \beta_1 x_i \]
\[ \alpha_{\text{SCHOOL}} \sim \text{Normal}(\alpha, \sigma) \]
\[ \beta_1 \sim \text{Normal}(0, 2) \]
\[ \alpha \sim \text{Normal}(0, 1) \]
\[ \sigma \sim \text{HalfCauchy}(0, 1) \]

Three additional models (Model I, Model II, and Model III) are proposed:

**Model I**
\[ \text{score}_i \sim \text{Normal}(\mu_i, p_i) \]
\[ \mu_i = \alpha + \alpha_{\text{SCHOOL}[i]} + \beta_1 x_i \]
\[ \alpha \sim \text{Normal}(0, 1) \]
\[ \alpha_{\text{SCHOOL}} \sim \text{Normal}(0, \sigma) \]
\[ \beta_1 \sim \text{Normal}(0, 2) \]
\[ \sigma \sim \text{HalfCauchy}(0, 1) \]

**Model II**
\[ \text{score}_i \sim \text{Normal}(\mu_i, p_i) \]
\[ \mu_i = \alpha_{\text{SCHOOL}[i]} + \beta_1 x_i \]
\[ \alpha_{\text{SCHOOL}} \sim \text{Normal}(\alpha, 2\sigma) \]
\[ \beta_1 \sim \text{Normal}(0, 2) \]
\[ \alpha \sim \text{Normal}(0, 1) \]
\[ \sigma \sim \text{HalfCauchy}(0, 1) \]

**Model III**
\[ \text{score}_i \sim \text{Normal}(\mu_i, p_i) \]
\[ \mu_i = \alpha_{\text{SCHOOL}[i]} + \beta_1 x_i \]
\[ \alpha_{\text{SCHOOL}} \sim \text{Normal}(0, 1) \]
\[ \beta_1 \sim \text{Normal}(0, 2) \]

Determine which of the three proposed models imposes additional model shrinkage:

A. None  
B. Model I and Model II only  
C. Model I and Model III only  
D. Model II and Model III only  
E. The answer is not given by (A), (B), (C), or (D)
29.

You are told that $\theta$ represents the probability of a positive outcome for the random variable $X$. Accordingly, you specify the following model:

$$X \sim \text{Bernoulli}(\theta)$$

$$p(\theta) \sim 1/\theta^2; \theta > 0$$

You observe $X$ for one trial and note that the outcome is positive. You select the following proposal distribution for the Metropolis algorithm:

$$\theta_{\text{prop}} = \begin{cases} 
\theta_{\text{curr}} + 0.1 & ; \text{probability } = 0.5 \\
\theta_{\text{curr}} - 0.1 & ; \text{probability } = 0.5 
\end{cases}$$

$\theta_{\text{prop}}$ and $\theta_{\text{curr}}$ represent the proposed and current parameter values.

The algorithm is seeded with an initial value for $\theta$ of 0.45.

Calculate the probability that after two iterations of the algorithm, $\theta$ is at 0.45.

A. Less than 0.15  
B. Greater than or equal to 0.15 but less than 0.25  
C. Greater than or equal to 0.25 but less than 0.35  
D. Greater than or equal to 0.35 but less than 0.45  
E. Greater than or equal to 0.45
30.

A parameter’s posterior is sampled using an MCMC algorithm. After discarding the warmup iterations, the first 6 samples from two chains from the sampler are provided in the order they were sampled:

chain 1 = {80, 100, 90, 90, 100, 110}
chain 2 = {95, 90, 85, 80, 70, 90}

Compute the Gelman-Rubin Convergence Diagnostic, $\hat{R}$, of the sampler.

A. Less than 0.85
B. Greater than or equal to 0.85 but less than 0.95
C. Greater than or equal to 0.95 but less than 1.05
D. Greater than or equal to 1.05 but less than 1.15
E. Greater than or equal to 1.15
31.

A dotchart of four models fit to the same underlying dataset is provided below. The dots represent estimations of in-sample and out-of-sample deviance for each of the four models as estimated by WAIC.

Determine which model has the most effective number of parameters.

A. model_A  
B. model_B  
C. model_C  
D. model_D  
E. All four models have the same number of effective number of parameters
The modeling team is building a new predictive model. While examining their input data, they produce the following scatterplot for analyzing available predictor variables.

The team originally fit a model with var2 and var4. The team decides to also include var3 in a refit model.

Identify which of following is the most likely to occur in the refit model.

A. Model predictions will be inaccurate
B. Standard errors for var2 will increase
C. Coefficient estimate for var3 will be negative
D. Posterior estimates will improve
E. Interpretation of the coefficients will be straightforward
A coin is tossed 10 times and 7 heads are observed. $\theta$ denotes the unknown probability of heads. You are given the following prior distribution for $\theta$:

$$p(\theta) = 2\cos^2(4\pi\theta)$$

where $\pi$ radians $= 180^\circ$

A Metropolis algorithm with a symmetric proposal distribution is being used to sample from the posterior distribution of $\theta$. Assume the $i^{th}$ draw from the posterior distribution has resulted in the value of 0.71. The $(i+1)^{th}$ draw is 0.70.

Determine the acceptance probability of the $(i+1)^{th}$ draw.

A. Less than 0.7  
B. Greater than or equal to 0.7 but less than 0.8  
C. Greater than or equal to 0.8 but less than 0.9  
D. Greater than or equal to 0.9 but less than 1  
E. 1.0
34.

Assume our data \(x_1, x_2, \ldots, x_n\) come from a normal distribution with an unknown mean \(\mu\) and known standard deviation, \(\sigma\). We assume a flat prior for \(\mu\).

We use the Metropolis algorithm to sample from the posterior distribution of \(\mu\) using a standard normal distribution as a proposal distribution. The chain is allowed to run for 3,000 iterations.

Calculate the probability that the 1000\(^{th}\) draw is greater than the 999\(^{th}\) draw.

A. Less than 0.30
B. Greater than or equal to 0.3 but less than 0.45
C. Greater than or equal to 0.45 but less than 0.60
D. Greater than or equal to 0.60 but less than 0.75
E. Greater than or equal to 0.75
35.

You assume that the number of claims for a single policy, $Y$, follows a Poisson distribution with mean parameter $\lambda$.

$$f(Y|\lambda) = \frac{e^{-\lambda} \lambda^Y}{Y!} \quad for \quad Y \in \{0, 1, 2, 3, \ldots\}$$

We initially choose an exponential prior for $\lambda$ with mean 7. You find that the mean number of claims per policy in your dataset is 2, and the mode of the posterior distribution of $\lambda$ is 2.1.

You are considering 3 possible alternative prior distributions.

I. Gamma with a mean of 7 and variance of 7
II. Exponential with a mean of 5
III. Uniform between 0 and 2

Determine which of the above alternative prior distributions will increase the mode (relative to the current mode) of the posterior distribution of $\lambda$.

A. None
B. I and II only
C. I and III only
D. II and III only
E. The answer is not given by (A), (B), (C), or (D)
You fit a model using MCMC and reviewed the results. You are presented with three potential problems and plausible solutions for those potential problems.

I. Potential Problem: Effective sample size is small as a result of autocorrelation of samples.  
Plausible Solution: Increase the number of iterations and thin the chain.

II. Potential Problem: The sampling algorithm has not adapted.  
Plausible Solution: Reduce the number of warmup iterations per chain.

III. Potential Problem: Your chain has converged but you are not certain that it converged in the correct region of the parameter space.  
Plausible Solution: Increase the number of iterations.

Determine which of the preceding represents both a potential problem and an appropriate plausible solution to that problem.

A. I only  
B. II only  
C. III only  
D. I, II and III  
E. The answer is not given by (A), (B), (C), or (D)
An actuary is comparing the predictions from two different Bayesian models, \( q \) and \( r \). Model \( q \) has 4 effective parameters and model \( r \) has 5 effective parameters. Below are the posterior likelihoods from the two models:

<table>
<thead>
<tr>
<th>Posterior Prediction</th>
<th>Posterior Likelihood from ( q )</th>
<th>Posterior Likelihood from ( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>150</td>
<td>175</td>
</tr>
<tr>
<td>3</td>
<td>175</td>
<td>200</td>
</tr>
<tr>
<td>4</td>
<td>125</td>
<td>125</td>
</tr>
</tbody>
</table>

The actuary decides to calculate Deviance and AIC and uses the most appropriate measure to identify the better model considering the effect of overfitting.

Identify which of the following statements is correct.

A. Model \( q \) has a better AIC, so it is a better model  
B. Model \( r \) has a better Deviance, so it is a better model  
C. Model \( q \) has a better AIC and Deviance, so it is a better model  
D. Model \( r \) has a better AIC and Deviance, so it is a better model  
E. Neither Deviance nor AIC measures overfitting, so the analysis is not conclusive
Given the following three statements about tree-based methods for regression and classification:

I. The main difference between bagging and random forests is the number of predictors used to build individual trees.
II. Single decision tree models generally have higher variance than random forest models.
III. Random forests provide an improvement over bagging because trees in a random forest are less correlated than those in bagged trees.

Determine which of the statements I, II, and III are true.

A. I only
B. II only
C. III only
D. I, II, and III
E. The answer is not given by (A), (B), (C), or (D)
Exam MAS-II Fall 2019

39.

Dataset $Z$ contains 4 variables and 100 records and has the following correlation matrix.

$$
\begin{bmatrix}
1.00 & 0.93 & 0.08 & -1.00 \\
0.93 & 1.00 & 0.10 & -0.93 \\
0.08 & 0.10 & 1.00 & -0.08 \\
-1.00 & -0.93 & -0.08 & 1.00 \\
\end{bmatrix}
$$

Determine which of the following plots of cumulative proportional variance is produced by principal components analysis on Dataset $Z$.

A. 

![Graph A]

B. 

![Graph B]

C. 

![Graph C]

D. 

![Graph D]

E. 

![Graph E]
40.

You are given three statements about the k-means clustering algorithm.

I. The k-means clustering algorithm requires that observations be standardized to have mean zero and standard deviation one.
II. The k-means clustering algorithm seeks to find subgroups of homogeneous observations.
III. The k-means clustering algorithm looks for a low-dimensional representation of the observations that explains a significant amount of the variance.

Determine which of the statements I, II or III are true.

A. I only
B. II only
C. III only
D. I, II, and III
E. The answer is not given by (A), (B), (C), or (D)
41.

You are given:

- A data set contains 500 observations for five predictor variables \((X_1, X_2, X_3, X_4, X_5)\).
- Each predictor has been standardized to have mean 0 and standard deviation 1.
- Each of the 500 observations takes the form \((x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i})\) for \(i\) ranging from 1 to 500.
- For each observation, a new predictor \(Z\) is calculated by projecting onto the first principal component.
- The projection for the \(i^{th}\) observation is denoted by \(z_i\).

The total variance present in the data set is equal to 4 and \(\sum_{i=1}^{500} z_i^2 = 750\).

Calculate the proportion of variance explained for the first principal component.

A. Less than or equal to 0.20
B. Greater than 0.20 but less than or equal to 0.30
C. Greater than 0.30 but less than or equal to 0.40
D. Greater than 0.40 but less than or equal to 0.50
E. Greater than 0.50
An actuary is using hierarchical clustering to group the following observations:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>40</td>
<td>60</td>
<td>71</td>
</tr>
</tbody>
</table>

The actuary recalculates the clustering using two linkage methods: complete and average.

- $h_{comp}$ is the height of the final fuse using complete linkage.
- $h_{avg}$ is the height of the final fuse using average linkage.

Calculate $|h_{comp} - h_{avg}|$.

A. Less than 10  
B. At least 10, but less than 15  
C. At least 15, but less than 20  
D. At least 20, but less than 25  
E. At least 25
<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td>22</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>E</td>
<td>23</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>E</td>
<td>24</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>25</td>
<td>D</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>26</td>
<td>B</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>27</td>
<td>C</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>28</td>
<td>A</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>29</td>
<td>D</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>30</td>
<td>E</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>31</td>
<td>A</td>
</tr>
<tr>
<td>11</td>
<td>C</td>
<td>32</td>
<td>B</td>
</tr>
<tr>
<td>12</td>
<td>E</td>
<td>33</td>
<td>C</td>
</tr>
<tr>
<td>13</td>
<td>D</td>
<td>34</td>
<td>A,B,C</td>
</tr>
<tr>
<td>14</td>
<td>B</td>
<td>35</td>
<td>E</td>
</tr>
<tr>
<td>15</td>
<td>A</td>
<td>36</td>
<td>A</td>
</tr>
<tr>
<td>16</td>
<td>B</td>
<td>37</td>
<td>A</td>
</tr>
<tr>
<td>17</td>
<td>D</td>
<td>38</td>
<td>D,E</td>
</tr>
<tr>
<td>18</td>
<td>D</td>
<td>39</td>
<td>C</td>
</tr>
<tr>
<td>19</td>
<td>C</td>
<td>40</td>
<td>B</td>
</tr>
<tr>
<td>20</td>
<td>C</td>
<td>41</td>
<td>C</td>
</tr>
<tr>
<td>21</td>
<td>A</td>
<td>42</td>
<td>B</td>
</tr>
</tbody>
</table>