Exam MAS-II

CASUALTY ACTUARIAL SOCIETY
AND THE
CANADIAN INSTITUTE OF ACTUARIES

Expertise. Insight. Solutions.

Canadian Institute of Actuaries / Institut canadien des actuaires

MAS-II

Modern Actuarial Statistics II

October 26, 2018

4 HOURS

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 2018 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

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, "Systolic Blood Pressure 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 9, 2018.

END OF INSTRUCTIONS

1.

In the supplemental material, you have been given a case study, "Systolic Blood Pressure Case Study", showing the results of different treatment options and the description of how that study was set up. There are different ways of setting up models to examine the benefits of the different treatment options. You have been asked which of two model structures will give a better fit to the experience.

Model Structure XYZ has:

- All eight treatment options in the fixed effects section of the model
- A random effect of doctors nested within hospitals
- An assumption of constant variance across treatment effects

Model Structure STW has:

- All eight treatment options in the fixed effects section of the model
- A random effect of doctors nested within hospitals
- An assumption that the variance by treatment can be grouped under Variance Group #1

The null hypothesis is that variance is constant across all treatment effects.

Determine the level of significance at which one would reject the null hypothesis using a likelihood ratio test.

A. Less than .005
B. At least .005, but less than .01
C. At least .01, but less than .025
D. At least .025, but less than .05
E. Greater than .05

CONTINUED ON NEXT PAGE

2.

In the supplemental material, you have been given a case study, "Systolic Blood Pressure Case Study", showing the results of different treatment options and the description of how that study was set up. There are different ways of setting up models to examine the benefits of the different treatment options. You have been asked which of two models structures will give a better fit to the experience.

Model Structure XYZ has:

- Treatment options should be grouped using Mean Group #1 in the fixed effects section of the model
- A random effect of doctors nested within hospitals
- An assumption that the variance by treatment can be grouped under Variance Group #1

Model Structure STW has:

- All eight treatment options in the fixed effects section of the model
- A random effect of doctors nested within hospitals
- An assumption that the variance by treatment can be grouped under Variance Group #1

The null hypothesis is that the Mean Group #1 should be retained to evaluate the effectiveness of treatment options.

Determine the level of significance at which one would reject the null hypothesis using a likelihood ratio test.

A. Less than .005
B. At least .005, but less than .01
C. At least .01, but less than .025
D. At least .025, but less than .05
E. Greater than .05

CONTINUED ON NEXT PAGE

3.

An insurance company sells homeowners' policies, each of which belongs to one of two possible risk groups, S and T. You are given the following information:

- Risk group S occurs 20% of the time.
- Risk group S has claim frequencies that are Poisson distributed with parameter $\lambda = 2$.
- Risk group S has claim severity that is uniformly distributed between 100 and 1000.

- Risk group T occurs 80% of the time.
- Risk group T has claim frequencies that are Poisson distributed with parameter $\lambda = 1$.
- Risk group T has claim severity that is uniformly distributed between 2000 and 8000.

- 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 three prior years of loss experience for a given risk.

Calculate the Bühlmann credibility factor for this risk.

A. Less than 0.17
B. At least 0.17, but less than 0.20
C. At least 0.20, but less than 0.23
D. At least 0.23, but less than 0.26
E. At least 0.26

CONTINUED ON NEXT PAGE

4.

An insurance company writes automobile policies in various regions across the country. You are given the following information:

- The company wrote 2,500,000 policies countrywide with a pure premium of 960.
- In Region X, this company wrote 72,000 policies with a pure premium of 530.
- The expected variance of the pure premium within each region is 1,800,000,000.
- The variance of the region pure premium means is 45,000.
- Within each region, the losses for each automobile policy are identically distributed.

Calculate the credibility-weighted pure premium for Region X using Bühlmann credibility.

A. Less than 610
B. At least 610, but less than 640
C. At least 640, but less than 670
D. At least 670, but less than 700
E. At least 700

CONTINUED ON NEXT PAGE

5.

You are given:

- The number of claims incurred per year for each insured has mean θ and variance 2θ.
- θ is uniformly distributed from 0.05 to 0.25.
- The following table of claim experience for a company:

Year	Number of Insureds	Number of Claims
1	200	15
2	250	20
3	225	25
4	250	---

Calculate the estimated claim count for this company in Year 4 using the Bühlmann-Straub credibility approach.

A. Less than 21
B. At least 21, but less than 23
C. At least 23, but less than 25
D. At least 25, but less than 27
E. At least 27

CONTINUED ON NEXT PAGE

6.

You are given the following information:

- A block of insurance policies had 1,384 claims this period.
- The claims had a mean loss of 55 and variance of loss of 6,010.
- The mean frequency of these claims is 0.085 per policy.
- The block has 21,000 policies.
- Full credibility is based on a coverage probability of 98% for a range of within 5% deviation from the true mean.

You calculate the partial-credibility factor for severity, Z_X, and the partial-credibility factor for pure premium, Z_P, using the limited-fluctuation credibility method.

Calculate the absolute difference between Z_X and Z_P.

A. Less than 0.05
B. At least 0.05, but less than 0.15
C. At least 0.15, but less than 0.25
D. At least 0.25, but less than 0.35
E. At least 0.35

CONTINUED ON NEXT PAGE

7.

You are building a Linear Mixed Model with longitudinal observations. The observations within a subject exhibit positive auto correlation. The variance of the residuals is 0.647.

Determine which of the following covariance matrices exhibits the first-order autoregressive structure.

A. $\begin{pmatrix} 0.647 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.647 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.647 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.647 \end{pmatrix}$

B. $\begin{pmatrix} 0.647 & 0.541 & 0.000 & 0.000 \\ 0.541 & 0.647 & 0.541 & 0.000 \\ 0.000 & 0.541 & 0.647 & 0.541 \\ 0.000 & 0.000 & 0.541 & 0.647 \end{pmatrix}$

C. $\begin{pmatrix} 0.793 & 0.146 & 0.146 & 0.146 \\ 0.146 & 0.793 & 0.146 & 0.146 \\ 0.146 & 0.146 & 0.793 & 0.146 \\ 0.146 & 0.146 & 0.146 & 0.793 \end{pmatrix}$

D. $\begin{pmatrix} 0.647 & 0.391 & 0.102 & 0.037 \\ 0.391 & 0.647 & 0.391 & 0.102 \\ 0.102 & 0.391 & 0.647 & 0.391 \\ 0.037 & 0.102 & 0.391 & 0.647 \end{pmatrix}$

E. $\begin{pmatrix} 0.647 & 0.518 & 0.415 & 0.332 \\ 0.518 & 0.647 & 0.518 & 0.415 \\ 0.415 & 0.518 & 0.647 & 0.518 \\ 0.332 & 0.415 & 0.518 & 0.647 \end{pmatrix}$

CONTINUED ON NEXT PAGE

8.

You are fitting a Linear Mixed Model to a set of 50 observations. Your model is of the form $Y_i = X_i\beta + Z_i u_i + \varepsilon_i$.

Your model includes 3 fixed covariates and 2 random covariates. The variance-covariance matrix of u_i is D.

Determine the dimensions of the matrix D.

A. 2 rows and 2 columns
B. 3 rows and 3 columns
C. 5 rows and 5 columns
D. 50 rows and 50 columns
E. The answer is not given by (A), (B), (C), or (D)

CONTINUED ON NEXT PAGE

9.

While fitting a Linear Mixed Model, problems arose with the estimation of the covariance parameters. The following alternative approaches for the estimation of the covariance parameters were analyzed:

I. Fit the implied marginal model.
II. Fit the marginal model with an unstructured covariance matrix.
III. Choose alternative starting values for covariance parameter estimates.

Determine which of the alternative approaches above are appropriate for the estimation of the covariance parameters.

A. None are appropriate
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)

CONTINUED ON NEXT PAGE

10.

Determine which of the following tests is not an appropriate alternative to the likelihood ratio tests of hypothesis for the parameters in a Linear Mixed Model.

A. A t-test to test hypotheses about a single fixed-effect parameter.
B. A Type I F-test to test for linear hypotheses about multiple fixed effects.
C. A Type III F-test to test for linear hypotheses about multiple fixed effects.
D. An omnibus Wald test to test linear hypotheses of the form $H_0\colon L\beta = 0$ vs. $H_A\colon L\beta \neq 0$.
E. A Wald z-test to test for covariance parameters.

CONTINUED ON NEXT PAGE

11.

You are given the following three statements regarding parameter estimates of Linear Mixed Models using maximum likelihood (ML) estimation and residual maximum likelihood (REML) estimation.

I. ML estimates of the covariance parameters are biased.
II. The variance of the ML estimates of the fixed effects are biased.
III. The variance of the REML estimates of the fixed effects are unbiased.

Determine which of the preceding statements are true.

A. None are 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)

CONTINUED ON NEXT PAGE

12.

Height measurements were taken on the same 90 individuals for three consecutive years and a Linear Mixed Model is built using Sex and Age to model the Height of each individual.

The following summary statistics are provided from the data:

Analysis Variable: Height						
Sex	**Age**	**Number of Observations**	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
F	12	45	156.89	4.32	148	166
F	13	45	158.96	4.48	148	168
F	14	45	160.84	5.19	148	173
M	12	45	152.58	4.97	138	163
M	13	45	159.29	7.64	144	177
M	14	45	166.20	8.55	147	181

Determine which of the following equations specifies the Linear Mixed Model for a given subject i measured at time t with the greatest number of possible random effects.

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

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

C. $HEIGHT_{ti} = \beta_0 + \beta_1 \times SEX_{ti} + \beta_2 \times AGE_{ti} + \beta_3 \times SEX_{ti} \times AGE_{ti}$
$+u_{0i} + u_{2i} \times AGE_{ti} + \varepsilon_{ti}$

D. $HEIGHT_{ti} = \beta_0 + \beta_1 \times SEX_{ti} + \beta_2 \times AGE_{ti} + \beta_3 \times SEX_{ti} \times AGE_{ti}$
$+u_{0i} + u_{2i} \times AGE_{ti} + u_{3i} \times SEX_{ti} \times AGE_{ti} + \varepsilon_{ti}$

E. $HEIGHT_{ti} = \beta_0 + \beta_1 \times SEX_{ti} + \beta_2 \times AGE_{ti} + \beta_3 \times SEX_{ti} \times AGE_{ti}$
$+u_{0i} + u_{1i} \times SEX_{ti} + u_{2i} \times AGE_{ti} + u_{3i} \times SEX_{ti} \times AGE_{ti} + \varepsilon_{ti}$

CONTINUED ON NEXT PAGE

13.

An actuary creates a Hierarchical Linear Mixed Model of patient outcomes, with random effects for clinics (Level 3) and doctors (Level 2). Doctors are clustered in clinics. The first clinic has two doctors and six patients, with patients assigned to doctors as shown in the table below.

The table provides a complete list of data for Clinic 1, which is the first among eight clinics studied.

Clinic ID	Doctor ID	Patient ID
1	1000	001
1	1000	002
1	1001	003
1	1001	004
1	1001	005
1	1001	006

The marginal variance-covariance matrix (rows and columns in order of patient ID) for Clinic 1 is as follows:

$$\begin{bmatrix} 132 & 27 & 11 & 11 & 11 & 11 \\ 27 & 132 & 11 & 11 & 11 & 11 \\ 11 & 11 & 132 & 27 & 27 & 27 \\ 11 & 11 & 27 & 132 & 27 & 27 \\ 11 & 11 & 27 & 27 & 132 & 27 \\ 11 & 11 & 27 & 27 & 27 & 132 \end{bmatrix}$$

Calculate the Intraclass Correlation Coefficient for patients with the same doctor.

A. Less than 0.1
B. At least 0.1, but less than 0.2
C. At least 0.2, but less than 0.3
D. At least 0.3, but less than 0.4
E. At least 0.4

CONTINUED ON NEXT PAGE

14.

You are given the following information:

Ten classrooms are randomly assigned to receive one of three new teaching methods, labeled R, S and T. There are 300 students in these 10 classrooms. Student test scores are measured before and after the new teaching methods are implemented.

The following variables are considered:

- CLASS_ID — Class ID number
- METHOD_R — Indicator Variable (teaching method R = 1)
- METHOD_S — Indicator Variable (teaching method S = 1)
- STUDENT_ID — Student ID number
- SEX — Indicator variable (female = 1)
- SCORE_PRE — Student's test score before the new teaching method is implemented
- SCORE — Student's test score after the new teaching method is implemented

You are given Model 1 for the score of STUDENT_ID *i* in CLASS_ID *j*.

Model	Model Specification
1	$SCORE_{ij} = \beta_0 + \beta_1 \times SEX_{ij} + \beta_2 \times SCORE\_PRE_{ij} + \beta_3 \times METHOD\_R_j + \beta_4 \times METHOD\_S_j + u_j + \varepsilon_{ij}$

- $u_j \sim N(0, \sigma_{class}^2)$
- $\varepsilon_{ij} \sim N(0, \sigma^2)$
- All u_j, ε_{ij} are mutually independent

Determine which of the following statements about Model 1 is true.

A. There are seven parameters in the model.
B. The value of β_1 varies across individual students.
C. σ_{class}^2 represents variance within the classroom.
D. σ^2 represents variance between the classrooms.
E. The intercept for each classroom is the same.

CONTINUED ON NEXT PAGE

15.

You are given Model 1 for the SCORE of STUDENT_ID i in CLASS_ID j.

Model	Model Specification
1	$SCORE_{ij} = \beta_0 + \beta_1 \times SEX_{ij} + \beta_2 \times SCORE\_PRE_{ij} + \beta_3 \times METHOD\_R_j + \beta_4 \times METHOD\_S_j + u_j + \varepsilon_{ij}$

The following assumptions apply to Model 1 and the models listed under A – E:

- $u_j \sim N(0, \sigma_{class}^2)$
- $\varepsilon_{ij} \sim N(0, \sigma^2)$
- *All* u_j, ε_{ij} are mutually independent

For hierarchical models:

- Level 1 is Student
- Level 2 is Classroom

Determine which of the following four hierarchical models is equivalent to Model 1.

A. Level 1 : $SCORE_{ij} = b_{0j} + \beta_1 \times SEX_{ij} + \beta_2 \times SCORE\_PRE_{ij} + u_j + \varepsilon_{ij}$
Level 2 : $b_{0j} = \beta_0 + \beta_3 \times METHOD\_R_j + \beta_4 \times METHOD\_S_j + u_j$

B. Level 1 : $SCORE_{ij} = b_{0j} + \beta_1 \times SEX_{ij} + \beta_2 \times SCORE\_PRE_{ij} + \varepsilon_{ij}$
Level 2 : $b_{0j} = \beta_0 + \beta_3 \times METHOD\_R_j + \beta_4 \times METHOD\_S_j + u_j$

C. Level 1 : $SCORE_{ij} = b_{0j} + \beta_1 \times SEX_{ij} + \beta_2 \times SCORE\_PRE_{ij} + \beta_3 \times METHOD\_R_j + \beta_4 \times METHOD\_S_j + \varepsilon_{ij}$
Level 2 : $b_{0j} = \beta_0 + u_j$

D. Level 1 : $SCORE_{ij} = b_{0j} + \beta_1 \times SEX_{ij} + \beta_2 \times SCORE\_PRE_{ij} + \beta_3 \times METHOD\_R_j + \beta_4 \times METHOD\_S_j + u_j$
Level 2 : $b_{0j} = \beta_0 + u_j$

E. None of (A), (B), (C), or (D) are equivalent to Model 1.

CONTINUED ON NEXT PAGE

16.

The Metropolis algorithm is chosen to sample the posterior distribution of a parameter θ. Assume the following:

- An unnormalized posterior defined by the function $f(\theta) = \begin{cases} 0, & \text{for } \theta < 0 \\ e^{-\frac{\theta}{2}}, & \text{for } \theta \geq 0 \end{cases}$
- A sampler initialization point of *3.0* before Iteration 1
- Each proposal in the Metropolis algorithm is the sum of the current position of the sampler and the step within a given iteration
- A proposal is accepted if the acceptance ratio, $\frac{f(\theta_{prop})}{f(\theta_{curr})}$, is greater than the random uniform number

The first four iterations for the sampler are provided in the table below.

Iteration	Step	Random Uniform Number
1	1.0	0.820
2	-1.4	0.939
3	2.2	0.233
4	-3.7	0.468

Calculate the position of the sampler after iteration 4.

A. Less than 1.0
B. At least 1.0, but less than 2.0
C. At least 2.0, but less than 3.0
D. At least 3.0, but less than 4.0
E. At least 4.0

CONTINUED ON NEXT PAGE

17.

A chain of parameters is sampled using the Gibbs algorithm. The length of the chain is 2,500. 25% of the samples are regarded as warm-up. The autocorrelation, ρ, is computed at each lag, k.

k	ρ_k
0	1.00
1	0.82
2	0.73
3	0.54
4	0.56
5	0.42
6	0.31
7	0.28
8	0.17
9	0.18
10	0.04
11	-0.01
12	0.02
13	-0.03
14	0.01
15	-0.03
16	0.02
17	0.00
18	0.05
19	-0.01
20	-0.02

Calculate the effective sample size of the chain.

A. Less than or equal to 200
B. Greater than 200 but less than or equal to 225
C. Greater than 225 but less than or equal to 250
D. Greater than 250 but less than or equal to 275
E. Greater than 275

CONTINUED ON NEXT PAGE

18.

You are given the following statements regarding Hamiltonian MCMC.

I. It is possible to use a relatively small number of effective posterior samples if you are only trying to measure the posterior mean.

II. It is always necessary to run multiple chains to make inferences from the posterior distribution.

III. In general, using weakly informative priors will improve the efficiency of the sampling compared to non-informative priors.

Determine which of the above statements are true.

A. None are 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)

CONTINUED ON NEXT PAGE

19.

The following statements about the Gelman-Rubin convergence diagnostic are provided.

I. It measures autocorrelation of samples within a chain.
II. It aids in determining whether there are a sufficient number of chains in the sampling process.
III. It is used to help assess convergence of multiple chains in the sampling process.

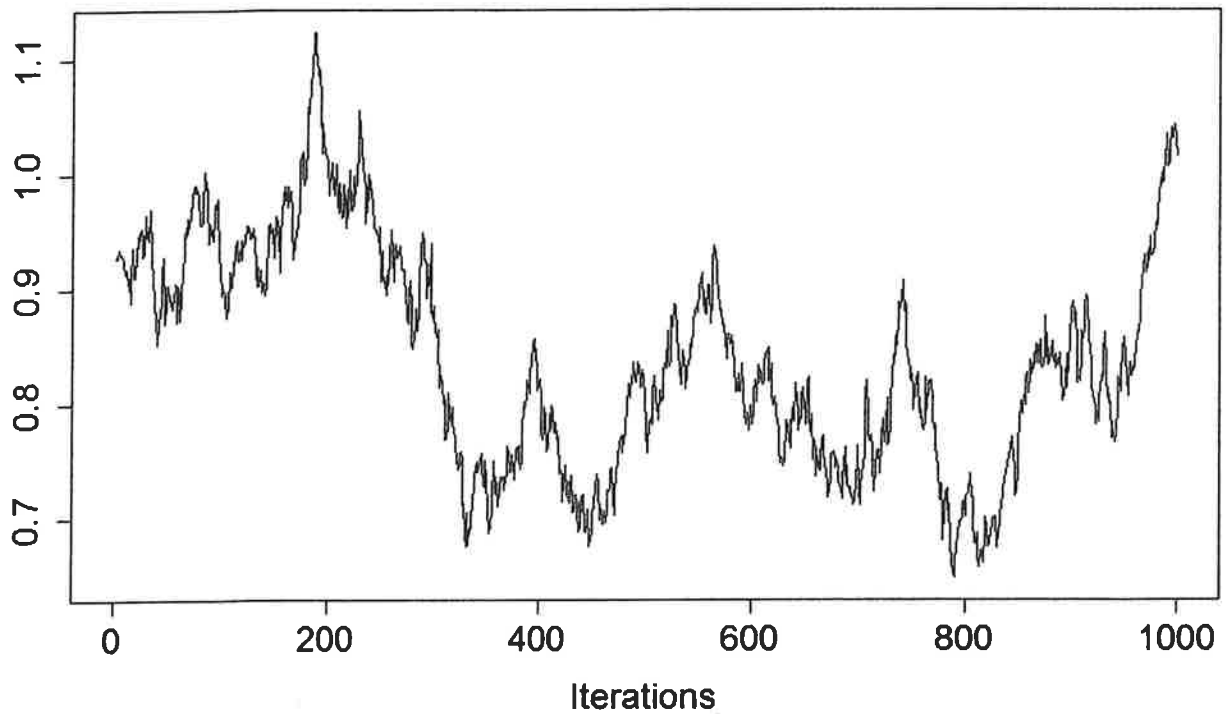
Determine which of the above statements are true.

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)

CONTINUED ON NEXT PAGE

20.

The trace plot for a Bayesian MCMC chain is provided below.



Determine which of the following best describes any problem with the chain and the action that should be taken.

A. There is a low level of auto-correlation; the mean of the likelihood distribution should be changed.
B. There is a low level of auto-correlation; many more iterations are needed to make inferences from the samples of the posterior.
C. There is a high level of auto-correlation; the variance of the prior distribution should be reduced.
D. There is a high level of auto-correlation; many more iterations are needed to make inferences from the samples of the posterior.
E. There are no issues with this chain.

CONTINUED ON NEXT PAGE

21.

In the table below, p_i represents the true distribution of events and q_i represents a model's estimates of p_i.

i	p_i	q_i
1	0.2	0.1
2	0.3	0.2
3	0.5	0.7

Calculate the Kullback-Leibler divergence.

A. Less than or equal to 0.090
B. Greater than 0.090 but less than or equal to 0.095
C. Greater than 0.095 but less than or equal to 0.100
D. Greater than 0.100 but less than or equal to 0.105
E. Greater than 0.105

CONTINUED ON NEXT PAGE

22.

You are given the following information:

Model	WAIC	Effective Number of Parameters (p_{WAIC})
M1	681.8	2.1
M2	682.4	3.0
M3	688.1	1.1

Calculate the Akaike weight for Model M2.

A. Less than 0.25
B. Greater than 0.25 but less than or equal to 0.35
C. Greater than 0.35 but less than or equal to 0.45
D. Greater than 0.45 but less than or equal to 0.55
E. Greater than 0.55

CONTINUED ON NEXT PAGE

23.

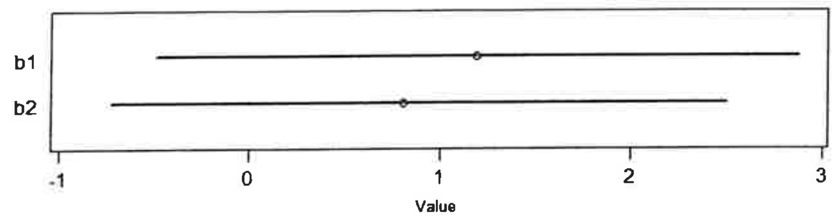
A linear regression is fit with the model below:

$$\begin{aligned} y_i &\sim Normal(\mu_i, \sigma) \\ \mu_i &= b_0 + b_1 x_{1i} + b_2 x_{2i} \\ b_0 &\sim Normal(10, 50) \\ b_1 &\sim Normal(2, 10) \\ b_2 &\sim Normal(2, 10) \\ \sigma &\sim Uniform(0, 10) \end{aligned}$$

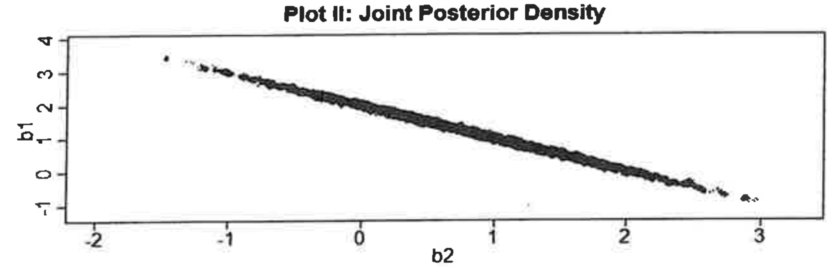
It is known that X_1 and X_2 are both highly correlated with Y.

Three posterior plots of b_1 and b_2 are provided below.

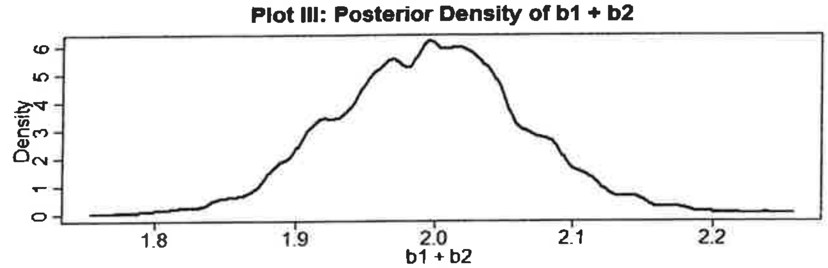
Plot I: Marginal Posterior Estimates and 95% Posterior Intervals



Plot II: Joint Posterior Density



Plot III: Posterior Density of b1 + b2



Determine the best interpretation regarding multicollinearity from these plots.

A. It is present in the data because x_1 and x_2 are highly positively correlated
B. It is present in the data because x_1 and x_2 are highly negatively correlated
C. It is present in the data because b_1 and b_2 are highly positively correlated
D. It is present in the data because x_1 and x_2 have low correlation with each other
E. It is not present in the data

CONTINUED ON NEXT PAGE

24.

The random variable Y follows a Poisson distribution with mean parameter λ. Six posterior samples for λ are provided in the table below:

i	λ_i
1	2
2	3
3	1
4	3
5	4
6	2

The following additional information is provided:

- The first realized observation from Y, y_1, is 3
- The contribution from y_1 to the effective number of parameters (p_{WAIC}) is 0.2019

Calculate the contribution to WAIC from y_1.

A. Less than 1.05
B. Greater than or equal to 1.05 but less than 2.05
C. Greater than or equal to 2.05 but less than 3.05
D. Greater than or equal to 3.05 but less than 4.05
E. Greater than or equal to 4.05

CONTINUED ON NEXT PAGE

25.

Consider the following statements about Deviance Information Criterion (DIC).

I. Higher values indicate better models.
II. DIC includes a penalty for the effective number of parameters.
III. DIC is similar to AIC except that it accounts for information in the prior distribution(s).

Determine which of the above statements are true.

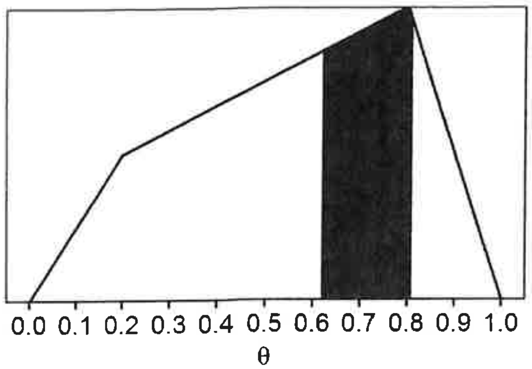
A. None are true
B. I and II only
C. I and III only
D. II and III only
E. The answer is not given by (A), (B), (C), or (D)

CONTINUED ON NEXT PAGE

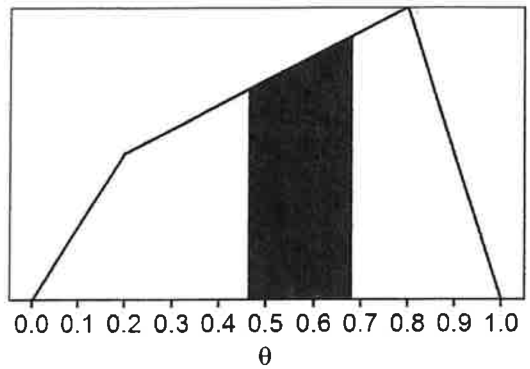
26.

The following charts present the posterior distribution of θ.

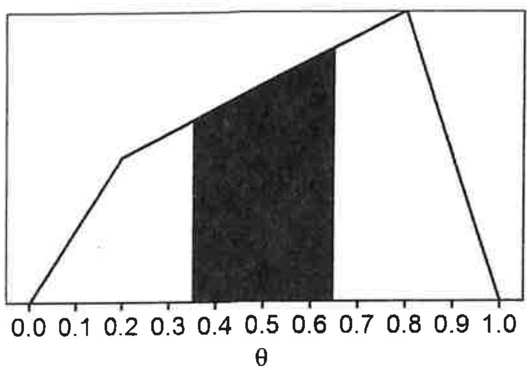
Plot 1



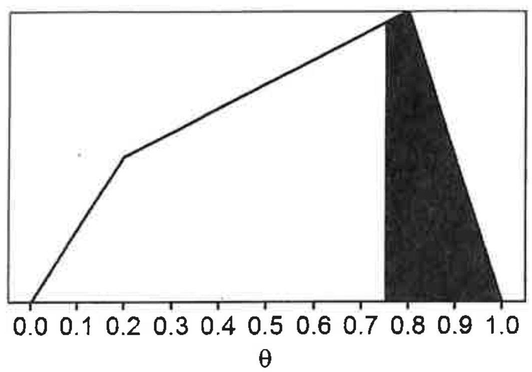
Plot 2



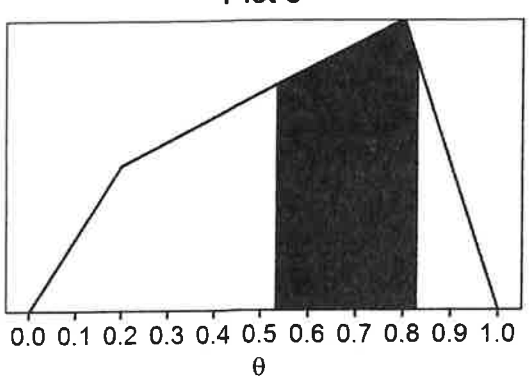
Plot 3



Plot 4



Plot 5



Determine which of the shaded regions in the above plots represents the 30% highest posterior density interval (HPDI).

A. Plot 1
B. Plot 2
C. Plot 3
D. Plot 4
E. Plot 5

CONTINUED ON NEXT PAGE

27.

A random variable X is Poisson distributed with mean parameter λ.

The posterior distribution of λ is:

$$\lambda = \begin{cases} 0.8; & probability = 0.55 \\ 1.2; & probability = 0.45 \end{cases}$$

A sample of size 10 is drawn from the posterior distribution of λ. The 10 sampled values are:
0.8, 0.8, 0.8, 1.2, 0.8, 1.2, 1.2, 0.8, 0.8, 1.2

The 10 sampled $\lambda's$ are stored in a vector named lambdas in R as follows:

```
lambdas <- c(0.8, 0.8, 0.8, 1.2, 0.8, 1.2, 1.2, 0.8, 0.8, 1.2)
```

A random sample of size 10 is drawn for the random variable X using the `lambdas` vector. The R function `rpois()` is used for sampling in the next line.

```
rpois(n = 10, lambda = lambdas)
```

The sampled values of X are:
1, 1, 3, 0, 1, 3, 0, 1, 1, 3

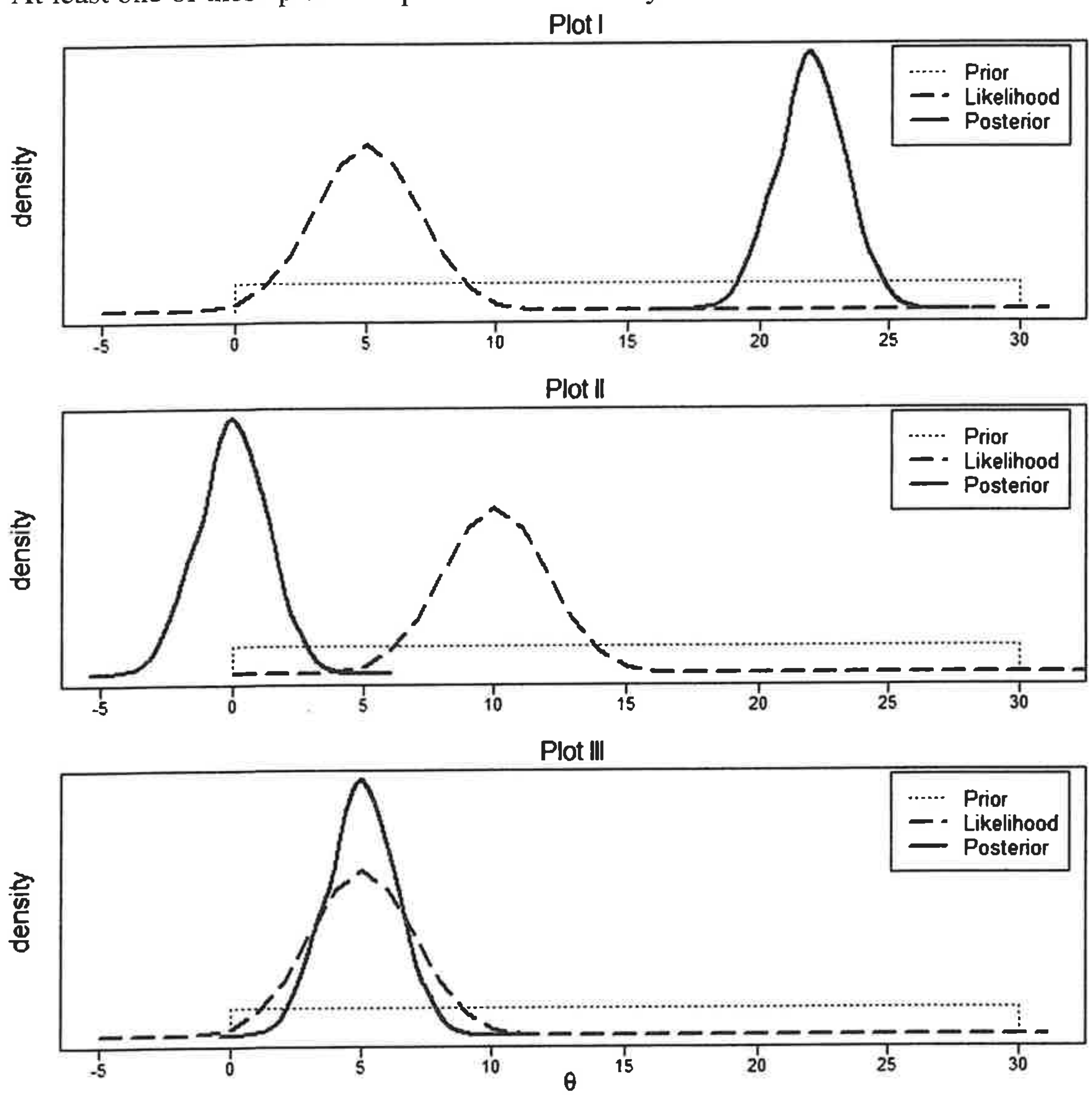
Calculate the posterior predictive probability that $X = 1$ using the sampling distribution above.

A. Less than or equal to 0.2
B. Greater than 0.2, but less than or equal to 0.4
C. Greater than 0.4, but less than or equal to 0.6
D. Greater than 0.6, but less than or equal to 0.8
E. Greater than 0.8

CONTINUED ON NEXT PAGE

28.

Three plots of likelihood, prior, and posterior densities for the parameter θ are provided below. At least one of these plots was produced from a Bayesian model.



Determine which of these three plots could have been produced by a Bayesian model.

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

CONTINUED ON NEXT PAGE

29.

You are given the following possible reasons to use a Bayesian multi-level model.

I. To improve estimates for repeat sampling
II. To improve estimates for imbalance in sampling
III. To avoid averaging

Determine which of the above reasons are valid.

A. None are valid
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)

CONTINUED ON NEXT PAGE

30.

You are given the following procedures:

I. Gibbs algorithm
II. Metropolis-Hastings algorithm
III. Restricted Maximum Likelihood

Determine which of the above procedures can be used to update posterior distributions.

A. None can be used
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)

CONTINUED ON NEXT PAGE

31.

You are given a vector of data, $\boldsymbol{y}$, a vector of parameters, $\boldsymbol{\theta}$, and a probability density function, p.

Determine which of the following expressions represents the likelihood function in a Bayesian model.

A. $p(\boldsymbol{y}|\boldsymbol{\theta})$
B. $p(\boldsymbol{\theta}|\boldsymbol{y})$
C. $p(\boldsymbol{y}, \boldsymbol{\theta})$
D. $p(\boldsymbol{y})$
E. $p(\boldsymbol{\theta})$

CONTINUED ON NEXT PAGE

32.

A linear model has the following form:

$$
\begin{array}{lcl}
Y_i & \sim & Normal(\mu_i,\ e^{\sigma}) \\
\mu_i & = & \theta_0 + \theta_1 X_{1i} + \theta_2 X_{1i}^2 + \theta_3 X_{2i} + \phi X_{3i} \\
\theta_0 & \sim & Normal(0,\ 10) \\
\theta_1 & \sim & Normal(0,\ 2) \\
\theta_2 & \sim & Normal(0,\ 0.2) \\
\theta_3 & \sim & Normal(-1,\ 1) \\
\phi & = & \gamma_1 + \gamma_2 X_{1i} \\
\gamma_1 & \sim & Normal(0,\ 3) \\
\gamma_2 & \sim & Normal(0.02,\ 0.6) \\
ln(\sigma) & \sim & Uniform(-3,\ 3)
\end{array}
$$

Determine the number of parameters in this model.

A. 5
B. 6
C. 7
D. 8
E. 9

CONTINUED ON NEXT PAGE

33.

You are given multiple different Bayesian linear models. To make a better prediction, you would like to consider all the provided models. To accomplish this, you are given the following options to evaluate:

I. Average the predictions of the multiple models
II. Average the parameter values of the multiple models
III. Discard and do not report on any but the best-performing model

Determine which of the above options are valid.

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

CONTINUED ON NEXT PAGE

34.

The Metropolis algorithm is used to estimate the posterior distribution of a parameter θ, which represents the probability of a flipped coin landing on heads.

Assume the following:

- The prior for θ is beta(1, 1)
- Two heads were observed in five flips
- Each proposed step in the Metropolis algorithm is the sum of the current position of the sampler and a draw from the proposal distribution

Three proposal distributions for the Metropolis algorithm are considered.

I. beta(2, 5)
II. uniform(0, 1)
III. uniform(current θ - 0.1, current θ + 0.1)

Determine which of the proposal distributions above are valid for this analysis.

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)

CONTINUED ON NEXT PAGE

35.

A linear regression is modeled with the following structure:

$$\begin{aligned} Y_i &\sim Poisson(\lambda_i) \\ ln(\lambda_i) &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \\ \beta_0 &\sim Normal(0,\ 5) \\ \beta_1 &\sim Normal(0,\ 2) \\ \beta_2 &\sim Normal(0,\ 2) \end{aligned}$$

100 samples from four posterior distributions are placed in order from smallest to largest. Each posterior is independently ordered. The i^{th} row corresponds to the i^{th} smallest sample.

i	β_0	β_1	β_2	$\beta_1 + \beta_2$
1	-5.59	-2.13	-2.86	-4.83
2	-4.89	-1.22	-2.40	-3.75
3	-4.68	-1.13	-2.37	-3.65
4	-4.54	-1.07	-1.92	-3.16
5	-4.28	-1.06	-1.89	-2.97
...	...	...	...	...
96	5.08	1.23	1.64	3.15
97	5.38	1.30	1.66	3.19
98	5.79	1.37	1.86	3.26
99	5.80	1.39	2.12	3.36
100	6.32	1.63	2.14	3.67

You are told that the 98^{th} posterior interval for β_0 is $[-4.89,\ 5.80]$.

Determine the upper bound of the 94^{th} posterior interval for $\beta_1 + \beta_2$.

A. Less than or equal to 3.00
B. Greater than 3.00 but less than or equal to 3.20
C. Greater than 3.20 but less than or equal to 3.40
D. Greater than 3.40 but less than or equal to 3.60
E. Greater than 3.60

CONTINUED ON NEXT PAGE

36.

A training data set contains eight observations for two predictor variables, X_1 and X_2, and a response variable, Y. The response Y has three possible classes P, N, and U.

i	X_1	X_2	Y	Distance from (x_{1i}, x_{2i}) to (3, 2)
1	4.1	3.0	P	1.5
2	-2.6	-3.0	N	7.5
3	-1.1	1.3	U	4.2
4	0.0	1.2	U	3.1
5	-3.0	-5.0	N	9.2
6	2.0	2.0	U	1.0
7	-3.1	-2.0	N	7.3
8	3.2	3.1	P	1.1

Three models are constructed using K-Nearest Neighbors and the data set above to predict Y for the two-dimensional space of predictors X_1 and X_2.

- Model I: K-Nearest Neighbors with K=1.
- Model II: K-Nearest Neighbors with K=3.
- Model III: K-Nearest Neighbors with K=7.

Each model is used to classify the point (3, 2).

Determine the predicted response Y at this point using each of the three models.

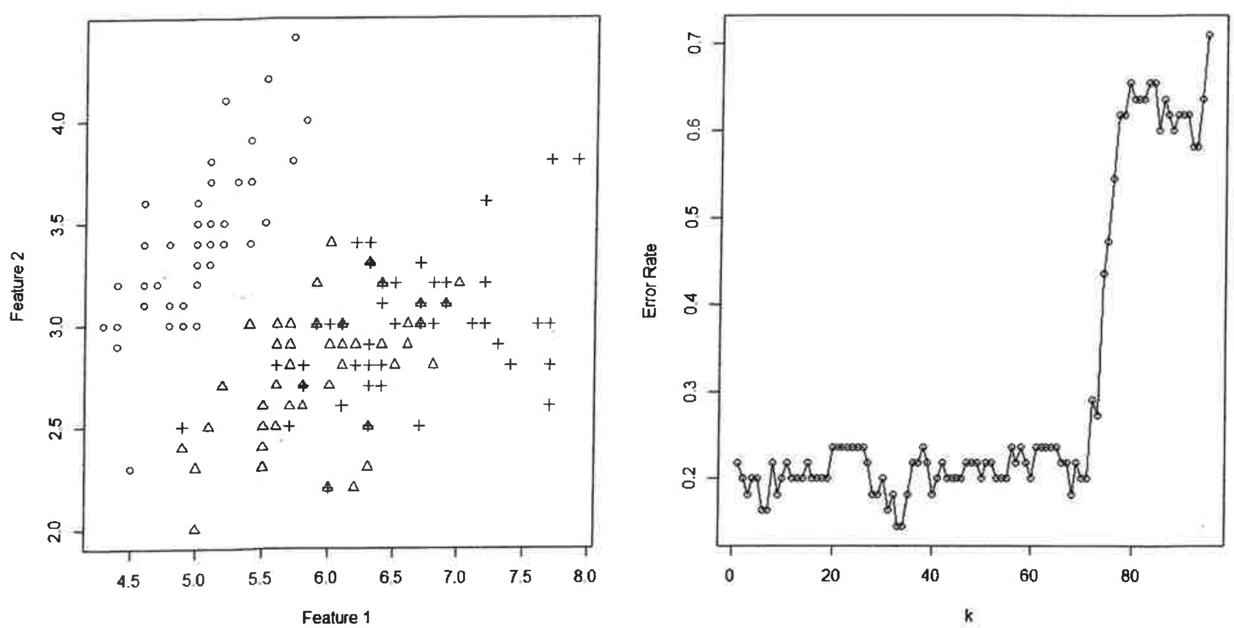
A. Model I: Y = P, Model II: Y = P, Model III: Y = P
B. Model I: Y = P, Model II: Y = P, Model III: Y = U
C. Model I: Y = U, Model II: Y = P, Model III: Y = P
D. Model I: Y = U, Model II: Y = P, Model III: Y = U
E. The answer is not given by (A), (B), (C), or (D).

CONTINUED ON NEXT PAGE

37.

You are given a data set consisting of 150 data points. There are three possible classifications for each data point, as shown in the leftmost graph, with 50 data points falling into each classification.

Based on this data you train a K-Nearest Neighbors model using 95 of the data points. You evaluate k ranging from 1 to 95. For each k, you calculate the test error rate on the remaining 55 data points with results shown in the rightmost graph.



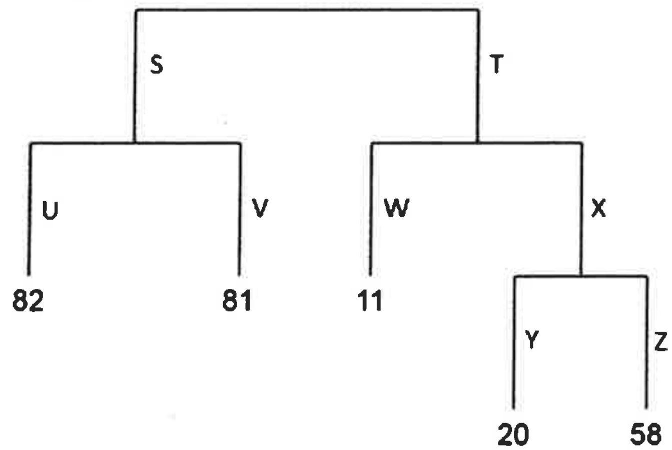
Determine the cause of the rapid increase in error rate between k equals 70 and k equals 95.

A. As k increases the K-Nearest Neighbors algorithm performs better.
B. As k increases the K-Nearest Neighbors algorithm performs worse.
C. As k approaches 95, all data points are predicted to have the same classification.
D. As k approaches 95, all data points are incorrectly classified.
E. There is no clear relationship between the value of k and the error rate.

CONTINUED ON NEXT PAGE

38:

You are given the following unpruned decision tree:



The values at each terminal node are the residual sums of squares (RSS) at that node. The table below gives the RSS at nodes S, T, and X if the tree was pruned at those nodes:

Node	RSS
S	251
T	209
X	86

The RSS for the null model is 486. You use the cost complexity pruning algorithm with the tuning parameter, α, equal to 9 in order to evaluate the following pruning strategies.

I. No nodes pruned
II. Prune node S only
III. Prune node T only
IV. Prune node X only
V. Prune both nodes S and X

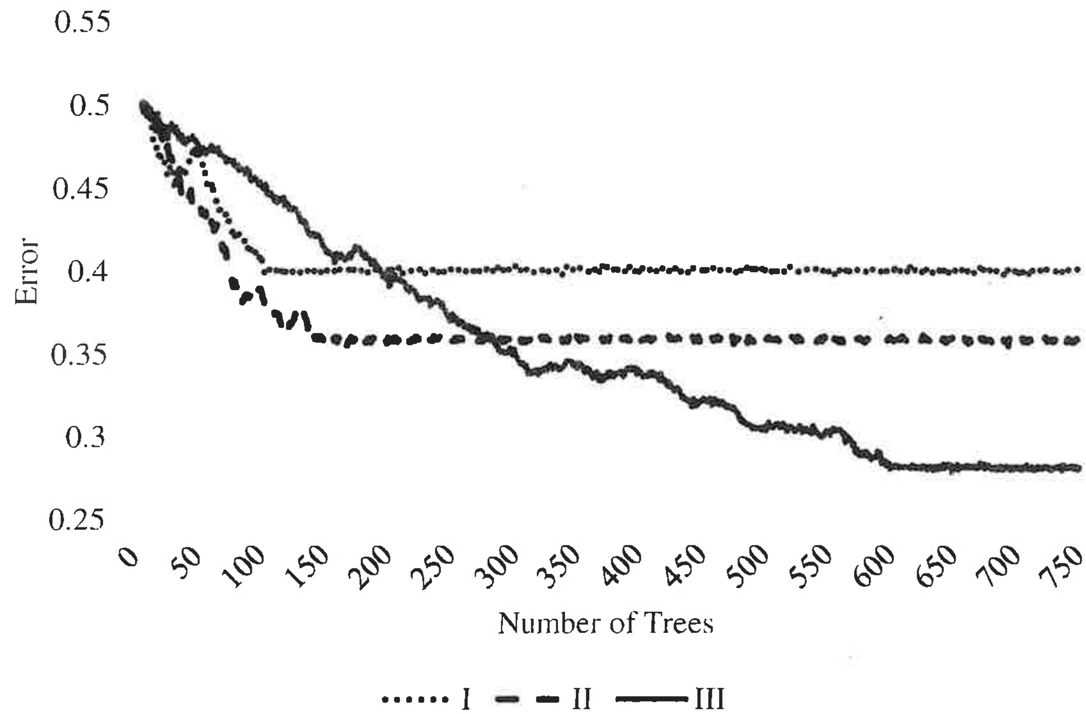
Determine which pruning strategy is selected.

A. I
B. II
C. III
D. IV
E. V

CONTINUED ON NEXT PAGE

39.

An actuary creates three tree-based models using bagging, boosting, and random forests. The error on a test data set, as a function of the number of trees in each model, is plotted on the graph below.



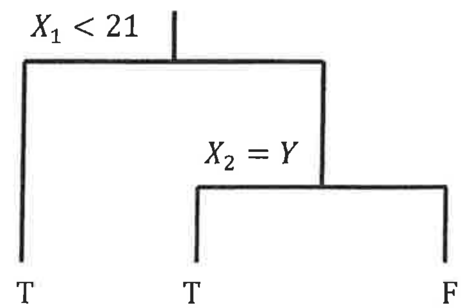
Determine the type of model most likely to have created each of the lines on the graph.

A. I: Boosting, II: Bagging, III: Random forest
B. I: Bagging, II: Boosting, III: Random forest
C. I: Bagging, II: Random forest, III: Boosting
D. I: Random forest, II: Bagging, III: Boosting
E. The answer is not given by (A), (B), (C), or (D).

CONTINUED ON NEXT PAGE

40.

You are given the following classification decision tree and data set:



i	X_1	X_2	Y
1	12	Y	T
2	23	N	F
3	4	Y	F
4	32	Y	F
5	22	N	T
6	30	Y	T
7	18	N	T

Determine the relationship between the classification error rate, the Gini index, and the cross-entropy, summed across all nodes.

A. cross-entropy > Gini index > classification error rate
B. cross-entropy > Gini index = classification error rate
C. classification error rate > Gini index > cross-entropy
D. Gini index > cross-entropy > classification error rate
E. The answer is not given by (A), (B), (C), or (D).

CONTINUED ON NEXT PAGE

41.

You are provided with the following normalized and scaled data set:

i	X_1	X_2	X_3
1	-0.577	1	-1
2	-0.577	1	1
3	-0.577	-1	1
4	1.732	-1	-1

The first principle component loading vector of the data set is (0.707, -0.500, -0.500).

Calculate the proportion of variance explained by the first principle component.

A. Less than 53%
B. At least 53% but less than 58%
C. At least 58% but less than 63%
D. At least 63% but less than 68%
E. At least 68%

CONTINUED ON NEXT PAGE

42.

You are provided the following data set with a single variable X.

i	X
1	9
2	15
3	4
4	2
5	18

A dendogram is built from this data set using agglomerative hierarchical clustering with complete linkage and Euclidean distance as the dissimilarity measure.

Calculate the tree height at which observation $i = 1$ fuses.

A. Less than 6
B. 6
C. 7
D. 8
E. At least 9

END OF EXAMINATION

FALL 2018 MAS-II
ANSWER KEY

	Answer		Answer
1	D	22	C
2	A	23	A
3	D	24	D
4	D	25	D
5	C	26	A
6	A	27	C
7	E	28	C
8	A	29	E
9	E	30	B
10	E	31	A
11	B	32	C
12	C	33	A
13	C	34	C & E
14	A	35	B
15	B	36	D
16	A	37	C
17	B	38	D
18	C	39	C
19	C	40	A
20	D	41	C
21	B	42	C