Exam MAS-I

CASUALTY ACTUARIAL SOCIETY
AND THE
CANADIAN INSTITUTE OF ACTUARIES

Jeanne Crowell
Vice President-Admissions

Jason Russ
Chairperson
Syllabus & Examination Committee

Frances Sarrel
General Officer-Syllabus
Syllabus & Examination Committee

MAS-I

Modern Actuarial Statistics I

October 25, 2018

4 HOURS

INSTRUCTIONS TO CANDIDATES

1. This 90 point examination consists of 45 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-I.
   - 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 "Tables for CAS MAS-I" 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, 2018.

END OF INSTRUCTIONS

1.

Insurance claims are made according to a Poisson process $\{N(t), t \geq 0\}$ with rate $\lambda = 1$.

Calculate $\mathrm{E}[N(1) * N(2)]$.

A. Less than 1.5
B. At least 1.5, but less than 2.5
C. At least 2.5, but less than 3.5
D. At least 3.5, but less than 4.5
E. At least 4.5

CONTINUED ON NEXT PAGE

2.

A Poisson process has the rate function $\lambda(t) = 1.5t$.

T_i is the time of the i^{th} event.

Calculate the probability that $T_2 > 3$.

A. Less than 1%
B. At least 1%, but less than 4%
C. At least 4%, but less than 7%
D. At least 7%, but less than 10%
E. At least 10%

CONTINUED ON NEXT PAGE

3.

You are given the following information regarding two portfolios of ABC Insurance Company:

- Claims in Portfolio 1 occur according to a Poisson process with a rate of four per year
- Claims in Portfolio 2 occur according to a Poisson process with a rate of two per year
- The two Poisson processes are independent

Calculate the probability that three claims occur in Portfolio 1 before three claims occur in Portfolio 2.

A. Less than 0.755
B. At least 0.755, but less than 0.765
C. At least 0.765, but less than 0.775
D. At least 0.775, but less than 0.785
E. At least 0.785

CONTINUED ON NEXT PAGE

4.

You are given the following information about a Poisson process:

- Claims occur at a rate of 4 per month
- Each claim is independent and takes on values with probabilities below:

Probability	Claim Amount
0.25	5
0.50	50
0.25	95

- The monthly claim experience is independent
- X is the aggregate claim amount for twelve months

Calculate Var[X].

A. Less than 40,000
B. At least 40,000, but less than 80,000
C. At least 80,000, but less than 120,000
D. At least 120,000, but less than 160,000
E. At least 160,000

CONTINUED ON NEXT PAGE

5.

You are given the following survival function for LED light bulbs in years:

$$S(x) = \left(1 - \frac{x}{30}\right)^{½}, \quad \text{for } 0 \leq x < 30$$

A homeowner replaced his kitchen light with an LED light bulb 24 months ago and it is still functioning today.

Calculate the probability that the LED light bulb will stop working between 24 months and 81 months from today.

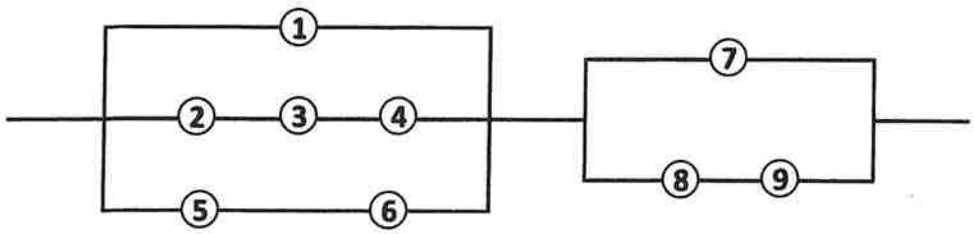
A. Less than 0.09
B. At least 0.09, but less than 0.10
C. At least 0.10, but less than 0.11
D. At least 0.11, but less than 0.12
E. At least 0.12

CONTINUED ON NEXT PAGE

6.

You are given:

- A system has 9 independent components, $i = 1, 2, 3, \dots, 9$
- $p_i = 0.75$, is the probability that the i^{th} component is functioning
- The system's structure is as pictured in the figure below:



- A new component with probability of functioning = 0.95 is available to replace one of the current components.
- The goal is to maximize the improvement of system's reliability.

Determine which one of the following components should be replaced to achieve the goal.

A. Component 1
B. Component 2
C. Component 5
D. Component 7
E. Component 8

CONTINUED ON NEXT PAGE

7.

A 3-out-of-50 system is placed in series with a 48-out-of-50-system.

Calculate the number of minimal path sets.

A. Fewer than 20,000
B. At least 20,000 but fewer than 100,000
C. At least 200,000 but fewer than 2,000,000
D. At least 2,000,000 but fewer than 20,000,000
E. At least 20,000,000

CONTINUED ON NEXT PAGE

8.

You are given the following information about a parallel system with two components:

- The first component has a lifetime that is uniform on (0, 1)
- The second component has a lifetime that is exponential with mean of 2

Determine which of the following is an expression for the expected lifetime of the system.

A. $\int_0^1 (1-t)dt + \int_0^1 t\, e^{-\frac{1}{2}t}dt + \int_1^\infty e^{-\frac{1}{2}t}dt$

B. $\int_0^\infty (1-t)e^{-\frac{1}{2}t}dt$

C. $\int_0^1 tdt + \int_0^1 (1-t)\,(1-e^{-\frac{1}{2}t})dt + \int_1^\infty (1-e^{-\frac{1}{2}t})dt$

D. $\int_0^\infty e^{-\frac{1}{2}t}dt + \int_0^1 (1-t)dt - \int_1^\infty (1-t)e^{-\frac{1}{2}t}dt$

E. $\int_0^1 (1-t)dt + \int_0^\infty e^{-\frac{1}{2}t}dt - \int_0^\infty (1-t)e^{-\frac{1}{2}t}dt$

CONTINUED ON NEXT PAGE

9.

You are given the following information about a homogenous Markov chain:

- There are three states to answering a trivia question: Skip (State 0), Correct (State 1), and Wrong (State 2)

- $P = \begin{bmatrix} 0 & 0.85 & 0.15 \\ 0.20 & 0.80 & 0 \\ 0 & 0.70 & 0.30 \end{bmatrix}$

- The candidate skipped the previous question

Calculate the probability of the candidate correctly answering at least one of the two subsequent questions.

A. Less than 0.92
B. At least 0.92, but less than 0.94
C. At least 0.94, but less than 0.96
D. At least 0.96, but less than 0.98
E. At least 0.98

CONTINUED ON NEXT PAGE

10.

You are given the following information about when a fair coin is flipped:

- If the outcome is Heads, 1 chip is won
- If the outcome is Tails, 1 chip is lost
- A gambler starts with 20 chips and will stop playing when he either has lost all his chips or he reaches 50 chips
- Of the first 10 flips, 7 are Heads and 3 are Tails

Calculate the probability that the gambler will lose all of his chips, given the results of the first 10 flips.

A. Less than 0.5
B. At least 0.5, but less than 0.6
C. At least 0.6, but less than 0.7
D. At least 0.7, but less than 0.8
E. At least 0.8

CONTINUED ON NEXT PAGE

11.

A scientist has discovered a way to create a new element. This scientist studied the new element and observed some of its properties:

- The element has a lifespan of 1 week, then it evaporates
- The element can produce offspring
- The creation of offspring is distributed as follows:

Number of Offspring	Probability
0	0.75
1	0.20
2	0.05

- The probability that the element will eventually die out is $\pi_0 = 1$

Determine which of the following best describes the above process.

A. Branching process
B. Gambler's ruin
C. Periodic Markov chain
D. Time reversible Markov chains
E. None of (A), (B), (C), or (D) describe this process

CONTINUED ON NEXT PAGE

12.

You are given the following information about a warranty insurance policy for a machine:

- The policy coverage allows for at most one claim
- The policy lasts for three years
- A benefit amount of 500 is paid at the end of year if a claim is made
- The annual probability that a machine is still functioning at the end of the year, given it was functioning at the beginning of the year, is as follows:

Year, t	p_t
0	0.90
1	0.80
2	0.70

- If there is no claim during the policy term, 100 is returned to the policyholder at the end of the policy term
- Annual interest rate, $i = 0.05$

Calculate the actuarial present value of this policy.

A. Less than 225
B. At least 225, but less than 250
C. At least 250, but less than 275
D. At least 275, but less than 300
E. At least 300

CONTINUED ON NEXT PAGE

13.

You are given the following information:

- There are two independent lives (35) and (55)
- The mortality of (35) follows the Illustrative Life Table
- The mortality of (55) follows the Illustrative Life Table, except that the annualized mortality after age 80 stays constant at the age 80 rate, where $q_{80} = 0.0803$

Calculate the probability that (35) is the only one of the two that lives to age 90.

A. Less than 0.075
B. At least 0.075, but less than 0.085
C. At least 0.085, but less than 0.095
D. At least 0.095, but less than 0.105
E. At least 0.105

CONTINUED ON NEXT PAGE

14.

An actuary is using the inversion method to simulate a random number from a distribution with the following density function:

$$f(x) = 5x^4, \qquad 0 < x < 1$$

A random draw of 0.6 was chosen from the uniform distribution $(0, 1)$.

Calculate the simulated random number.

A. Less than 0.2
B. At least 0.2, but less than 0.4
C. At least 0.4, but less than 0.6
D. At least 0.6, but less than 0.8
E. At least 0.8

CONTINUED ON NEXT PAGE

15.

You are given the following:

- A random variable, X, is uniformly distributed on the interval $(0, \theta)$
- θ is unknown
- For a random sample of size n an estimate of θ is given by:

$$\hat{\theta} = \frac{2}{n}\sum_{i=1}^{n} X_i$$

Calculate $\text{Var}(\hat{\theta})$.

A. $\frac{4\theta^2}{3n}$

B. $\frac{\theta^2}{3n^2}$

C. $\frac{\theta^2}{12}$

D. $\frac{\theta^2}{6}$

E. $\frac{\theta^2}{3n}$

CONTINUED ON NEXT PAGE

16.

Losses occur independently with the probabilities as described below:

Loss Size	Probability
0	25%
2	60%
6	15%

Using a sample of size 2, the variance of individual losses is estimated using the sample variance given by:

$$\sum_{i=1}^{2}(x_i - \bar{x})^2$$

Calculate the mean square error of this estimator.

A. Less than 8.0
B. At least 8.0, but less than 16.0
C. At least 16.0, but less than 24.0
D. At least 24.0, but less than 32.0
E. At least 32.0

CONTINUED ON NEXT PAGE

17.

Suppose that $X_1, \ldots, X_{10}$ is a random sample from a normal distribution $N(0, \sigma^2)$ such that:

$$\sum_{i=1}^{10} X_i = 10 \quad \text{and} \quad \sum_{i=1}^{10} X_i^2 = 500$$

Calculate the value of minimum variance unbiased estimator of σ^2.

A. Less than 30
B. At least 30, but less than 50
C. At least 50, but less than 70
D. At least 70, but less than 90
E. At least 90

CONTINUED ON NEXT PAGE

18.

An exponential distribution is fitted to the loss size data below using the method of moments.

Interval	Number of Observations
(0, 100]	5,000
(100, 500]	3,000
(500, 1000]	1,500
(1000, 5000]	500

Losses are assumed to follow a uniform distribution within each interval.

Calculate the fitted probability of a loss exceeding 500.

A. Less than 0.15
B. At least 0.15, but less than 0.25
C. At least 0.25, but less than 0.35
D. At least 0.35, but less than 0.45
E. At least 0.45

CONTINUED ON NEXT PAGE

19.

An actuary observes the following 20 losses for a select insurance policy:

$$\{20, 25, 36, 38, 42, 52, 55, 57, 65, 66, 69, 71, 72, 73, 74, 74, 74, 80, 81, 82\}$$

She believes the distribution which fits the data the best is loglogistic with the following probability density:

$$f(x;\gamma,\theta) = \frac{\gamma x^{\gamma-1}}{\theta^{\gamma}\left[1+\left(\frac{x}{\theta}\right)^{\gamma}\right]^{2}}$$

She uses the 20th and 80th percentiles to estimate the two parameters of this distribution.

Calculate the estimated value of θ.

A. Less than 20
B. At least 20, but less than 30
C. At least 30, but less than 40
D. At least 40, but less than 50
E. At least 50

CONTINUED ON NEXT PAGE

20.

X_i is the severity of claim i, which has an exponential distribution with mean = θ.

The payment for the claim under an insurance policy is capped at u.

There are $(n + s)$ total claims, with $\{x_1, x_2, \dots, x_n\}$ claims with payment less than u, and s claims with payment capped at u.

Determine which of the following is the MLE for θ.

A. $\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{u}{s}$

B. $\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{nu}{s}$

C. $\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{u}{n}$

D. $\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{su}{n}$

E. None of (A), (B), (C), or (D) are correct

CONTINUED ON NEXT PAGE

21.

Two independent populations X and Y have the following density functions:

- $f(x) = \lambda x^{\lambda-1}$ for $0 < x < 1$
- $g(y) = \mu y^{\mu-1}$ for $0 < y < 1$

$X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are random samples of sizes n and m, from X and Y, respectively.

You want to perform the following hypothesis test:

- $H_0: \lambda = 2; \mu = 4$
- $H_1: \lambda = 3; \mu = 8$

Determine the form of the best critical region for this test, using the Neyman-Pearson lemma.

A. $\sum_{i=1}^{n} \ln X_i + 4\sum_{i=1}^{m} \ln Y_i \leq c$

B. $\sum_{i=1}^{n} \ln X_i + 4\sum_{i=1}^{m} \ln Y_i \geq c$

C. $4\sum_{i=1}^{n} \ln X_i - \sum_{i=1}^{m} \ln Y_i \leq c$

D. $4\sum_{i=1}^{n} \ln X_i - \sum_{i=1}^{m} \ln Y_i \geq c$

E. The answer is not given by (A), (B), (C), or (D)

CONTINUED ON NEXT PAGE

22.

You are given the following table of test scores for four pairs of brothers and sisters:

Pair	1	2	3	4
Brother	89	94	78	80
Sister	86	95	71	77

- All test scores are normally distributed with a common variance
- Test scores between a brother and his sister are not independent
- You want to test the hypothesis:
    - H_0: Mean test scores of brother and sister are equal
    - H_1: Mean test scores of brother and sister are not equal

Calculate the p-value of this test.

A. Less than 0.01
B. At least 0.01, but less than 0.02
C. At least 0.02, but less than 0.05
D. At least 0.05, but less than 0.10
E. At least 0.10

CONTINUED ON NEXT PAGE

23.

A six-sided die is rolled 180 times and the following results were recorded:

Roll Result	1	2	3	4	5	6
Frequency	30	30	30	30	X	Y

- You use a Chi-squared test to evaluate the following hypothesis:
  - H_0: The die is fair (each roll result is equally likely)
  - H_1: The die is not fair
- The test significance level $\alpha = 0.05$
- $X < Y$

Calculate the largest value of X that would lead to a rejection of the null hypothesis.

A. Less than 17
B. 17
C. 18
D. 19
E. At least 20

CONTINUED ON NEXT PAGE

24.

The distribution of Y is given as:

$$F(y) = 1 - e^{-11y}, \quad \text{for } y > 0$$

Let $X = 5(e^{Y} - 1)$.

Calculate the mean of X.

A. At least 0.28, but less than 0.35
B. At least 0.35, but less than 0.42
C. At least 0.42, but less than 0.49
D. At least 0.49, but less than 0.56
E. At least 0.56

CONTINUED ON NEXT PAGE

25.

You draw a large number of independent samples, each of size $n = 4$, from a uniform distribution on $(0, \theta)$. You want to use the second smallest value in each sample as an estimate for the mean.

The density for the k[th] order statistic of a sample is given as:

$$g_k(y_k) = \frac{n!}{(k-1)!\,(n-k)!}[F(y_k)]^{k-1}[1-F(y_k)]^{n-k}f(y_k)$$

Calculate the expected bias of this estimate.

A. $-\frac{4\theta}{5}$

B. $-\frac{4\theta}{7}$

C. $-\frac{\theta}{10}$

D. $-\frac{\theta}{30}$

E. The answer is not given by (A), (B), (C), or (D)

CONTINUED ON NEXT PAGE

26.

You are given the following information from a model constructed to predict the probability that a Homeowners policy will be retained into the next policy term:

Response variable		retention
Response distribution		binomial
Link		probit
Pseudo R^2		0.6521
Parameter	df	$\hat{\beta}$
Intercept	1	0.4270
Tenure		
< 5 years	0	0.0000
≥ 5 years	1	0.1320
Prior Rate Change		
< 0%	1	0.0160
[0%, 10%]	0	0.0000
> 10%	1	-0.0920
Amount of Insurance (000's)	1	0.0015

Let $\hat{\pi}$ be the modeled probability that a policy with 4 years of tenure who experienced a +12% prior rate change and has 225,000 in amount of insurance will be retained into the next policy term.

Calculate $\hat{\pi}$.

A. Less than 0.60
B. At least 0.60, but less than 0.70
C. At least 0.70, but less than 0.80
D. At least 0.80, but less than 0.90
E. At least 0.90

CONTINUED ON NEXT PAGE

27.

You are given the following three functions of a random variable, y, where $-\infty < y < \infty$.

I. $g(y) = 2 + 3y + 3(y-5)^2$
II. $g(y) = 4 - 4y$
III. $g(y) = |y|$

Determine which of the above could be used as link functions in a GLM.

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

28.

In a study 100 subjects were asked to choose one of three election candidates (A, B or C). The subjects were organized into four age categories: (18-30, 31-45, 45-61, 61+).

A logistic regression was fitted to the subject responses to predict their preferred candidate, with age group (18-30) and Candidate A as the reference categories.

For age group (18-30), the log-odds for preference of Candidate B and Candidate C were -0.535 and -1.489 respectively.

Calculate the modeled probability of someone from age group (18-30) preferring Candidate B.

A. Less than 20%
B. At least 20%, but less than 40%
C. At least 40%, but less than 60%
D. At least 60%, but less than 80%
E. At least 80%

CONTINUED ON NEXT PAGE

29.

An ordinary least squares model with one variable (Advertising) and an intercept was fit to the following observed data in order to estimate Sales:

Observation	Advertising	Sales
1	5.5	100
2	5.8	110
3	6.0	112
4	5.9	115
5	6.2	117

Calculate the residual for the 3rd observation.

A. Less than -2
B. At least -2, but less than 0
C. At least 0, but less than 2
D. At least 2, but less than 4
E. At least 4

CONTINUED ON NEXT PAGE

30.

An actuary uses statistical software to run a regression of the median price of a house on 12 predictor variables plus an intercept. He obtains the following (partial) model output:

Residual standard error: 4.74 on 493 degrees of freedom
Multiple R-squared: 0.7406
F-statistic: 117.3 on 12 and 493 DF
p-value: < 2.2e-16

Calculate the adjusted R^2 for this model.

A. Less than 0.70
B. At least 0.70, but less than 0.72
C. At least 0.72, but less than 0.74
D. At least 0.74, but less than 0.76
E. At least 0.76

CONTINUED ON NEXT PAGE

31.

An actuary fits two GLMs, M_1 and M_2, to the same data in order to predict the probability of a customer purchasing an automobile insurance product. You are given the following information about each model:

Model	Explanatory Variables Included in Model	Degrees of Freedom Used	Log Likelihood
M_1	• Offered Price • Number of Vehicles • Age of Primary Insured • Prior Insurance Carrier	10	-11,565
M_2	• Offered Price • Number of Vehicles • Age of Primary Insured • Gender of Primary Insured • Credit Score of Primary Insured	8	-11,562

The actuary wants to evaluate which of the two models is superior.

Determine which of the following is the best course of action for the actuary to take.

A. Perform a likelihood ratio test
B. Compute the F-statistic and perform an F-test
C. Compute and compare the deviances of the two models
D. Compute and compare the AIC statistics of the two models
E. Compute the Chi-squared statistic and perform a Chi-squared test

CONTINUED ON NEXT PAGE

32.

An actuary uses a multiple regression model to estimate money spent on kitchen equipment using income, education, and savings. He uses 20 observations to perform the analysis and obtains the following output:

Coefficient	Estimate	Standard Error	t-value
Intercept	0.15085	0.73776	0.20447
Income	0.26528	0.10127	2.61953
Education	6.64357	2.01212	3.30178
Savings	7.31450	2.73977	2.66975

	Sum of Squares
Regression	2.65376
Total	7.62956

He wants to test the following hypothesis:

- $H_0: \beta_1 = \beta_2 = \beta_3 = 0$
- H_1: At least one of $\{\beta_1, \beta_2, \beta_3\}$ <> 0

Calculate the value of the F-statistics used in this test.

A. Less than 1
B. At least 1, but less than 3
C. At least 3, but less than 5
D. At least 5
E. The answer cannot be computed from the information given.

CONTINUED ON NEXT PAGE

33.

Two ordinary least squares models were built to predict expected annual losses on Homeowners policies. Information for the two models is provided below:

Model 1

Parameter	$\hat{\beta}$	p-value
Intercept	212	
Replacement Cost (000s)	0.03	< .001
Roof Size	0.15	< .001
Precipitation Index	120	0.02
Replacement Cost (000s) x Roof Size	0.0010	0.05

Model 2

Parameter	$\hat{\beta}$	p-value
Intercept	315	
Replacement Cost (000s)	0.02	< .001
Roof Size	0.17	< .001
Number of Bathrooms	210	0.03
Replacement Cost (000s) x Roof Size	0.0015	< .001

Model 1

Model Statistics	
R^2	0.91
Adj R^2	0.87
MSE	31,765
AIC	25,031

Model 2

Model Statistics	
R^2	0.94
Adj R^2	0.89
MSE	30,689
AIC	25,636

Model 1

Cross Validation Set	MSE
1	33,415
2	38,741
3	32,112
4	37,210
5	29,501

Model 2

Cross Validation Set	MSE
1	26,666
2	38,554
3	39,662
4	36,756
5	30,303

You use 5-fold cross validation to select superior of the two models.

Calculate the predicted expected annual loss for a homeowners policy with a 500,000 replacement cost, a 2,000 roof size, a 0.89 precipitation index and three bathrooms, using the selected model.

A. Less than 1,000
B. At least 1,000, but less than 1,500
C. At least 1,500, but less than 2,000
D. At least 2,000, but less than 2,500
E. At least 2,500

CONTINUED ON NEXT PAGE

34.

A least squares regression model is fit to a data set. The resulting predictors and their standard errors are shown below:

	Fitted	Standard Error
$\hat{\beta}_0$	1.00	0.05
$\hat{\beta}_1$	2.50	0.15
$\hat{\beta}_2$	-0.50	0.20
$\hat{\beta}_3$	2.00	0.80

A second model is fit to the same data using only the 3rd independent variable:

	Fitted	Standard Error
$\hat{\beta}_0$	1.25	0.08
$\hat{\beta}_3$	2.20	0.75

Calculate variance inflation factor of $\hat{\beta}_3$.

A. Less than 1.00
B. At least 1.00, but less than 1.05
C. At least 1.05, but less than 1.10
D. At least 1.10, but less than 1.15
E. At least 1.15

CONTINUED ON NEXT PAGE

35.

You are fitting a linear regression model of the form:

$$y = X\beta + e; \quad e_i \sim N(0, \sigma^2)$$

and are given the following values used in this model:

$$X = \begin{bmatrix} 1 & 0 & 1 & 9 \\ 1 & 1 & 1 & 15 \\ 1 & 1 & 1 & 8 \\ 1 & 1 & 0 & 7 \\ 1 & 1 & 0 & 6 \\ 1 & 0 & 0 & 6 \end{bmatrix}; \quad y = \begin{bmatrix} 19 \\ 32 \\ 19 \\ 17 \\ 13 \\ 15 \end{bmatrix}; \quad X^T X = \begin{bmatrix} 6 & 4 & 3 & 51 \\ 4 & 4 & 2 & 36 \\ 3 & 2 & 3 & 32 \\ 51 & 36 & 32 & 491 \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} 1.75 & -0.20 & 0.54 & -0.20 \\ -0.20 & 0.84 & 0.25 & -0.06 \\ 0.54 & 0.25 & 1.38 & -0.16 \\ -0.20 & -0.06 & -0.16 & 0.04 \end{bmatrix}; \quad (X^T X)^{-1} X^T y = \begin{bmatrix} 2.335 \\ 0.297 \\ -0.196 \\ 1.968 \end{bmatrix}$$

$$H = X(X^T X)^{-1} X^T = \begin{bmatrix} 0.684 & 0.070 & 0.247 & -0.171 & -0.146 & 0.316 \\ 0.070 & 0.975 & -0.044 & 0.108 & -0.038 & -0.070 \\ 0.247 & -0.044 & 0.797 & 0.063 & 0.184 & -0.247 \\ -0.171 & 0.108 & 0.063 & 0.418 & 0.411 & 0.171 \\ -0.146 & -0.038 & 0.184 & 0.411 & 0.443 & 0.146 \\ 0.316 & -0.070 & -0.247 & 0.171 & 0.146 & 0.684 \end{bmatrix}$$

Calculate the residual for the 5th observation.

A. Less than -1
B. At least -1, but less than 0
C. At least 0, but less than 1
D. At least 1, but less than 2
E. At least 2

CONTINUED ON NEXT PAGE

36.

A modeler creates a local regression model. After reviewing the results, the fitted line appears too wiggly, over-responding to trends in nearby data points. The modeler would like to adjust the model to produce more intuitive results.

Determine which one of the following adjustments the modeler should make.

A. Add a linear constraint in the regions before and after the first knot
B. Increase the number of orders in the regression equation
C. Increase the number of knots in the model
D. Reduce the number of knots in the model
E. Increase the span, s, of the model

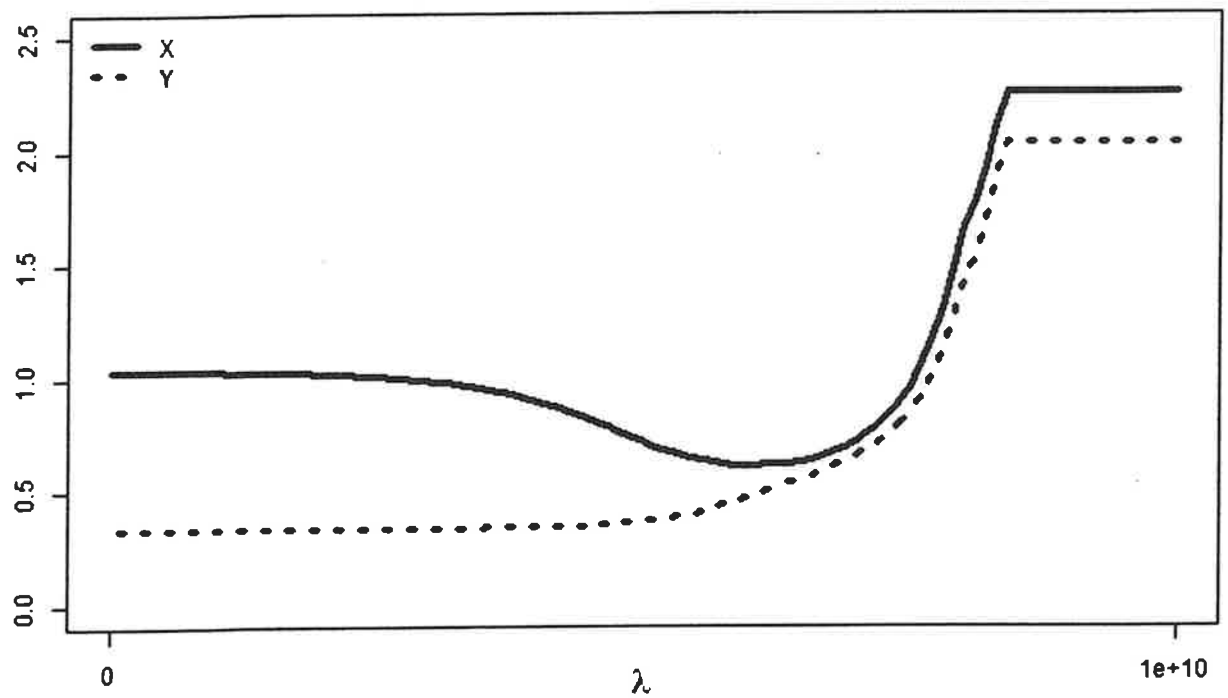
CONTINUED ON NEXT PAGE

37.

You are estimating the coefficients of a linear regression model by minimizing the sum:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

From this model you have produced the following plot of various statistics as a function of tuning parameters, λ:



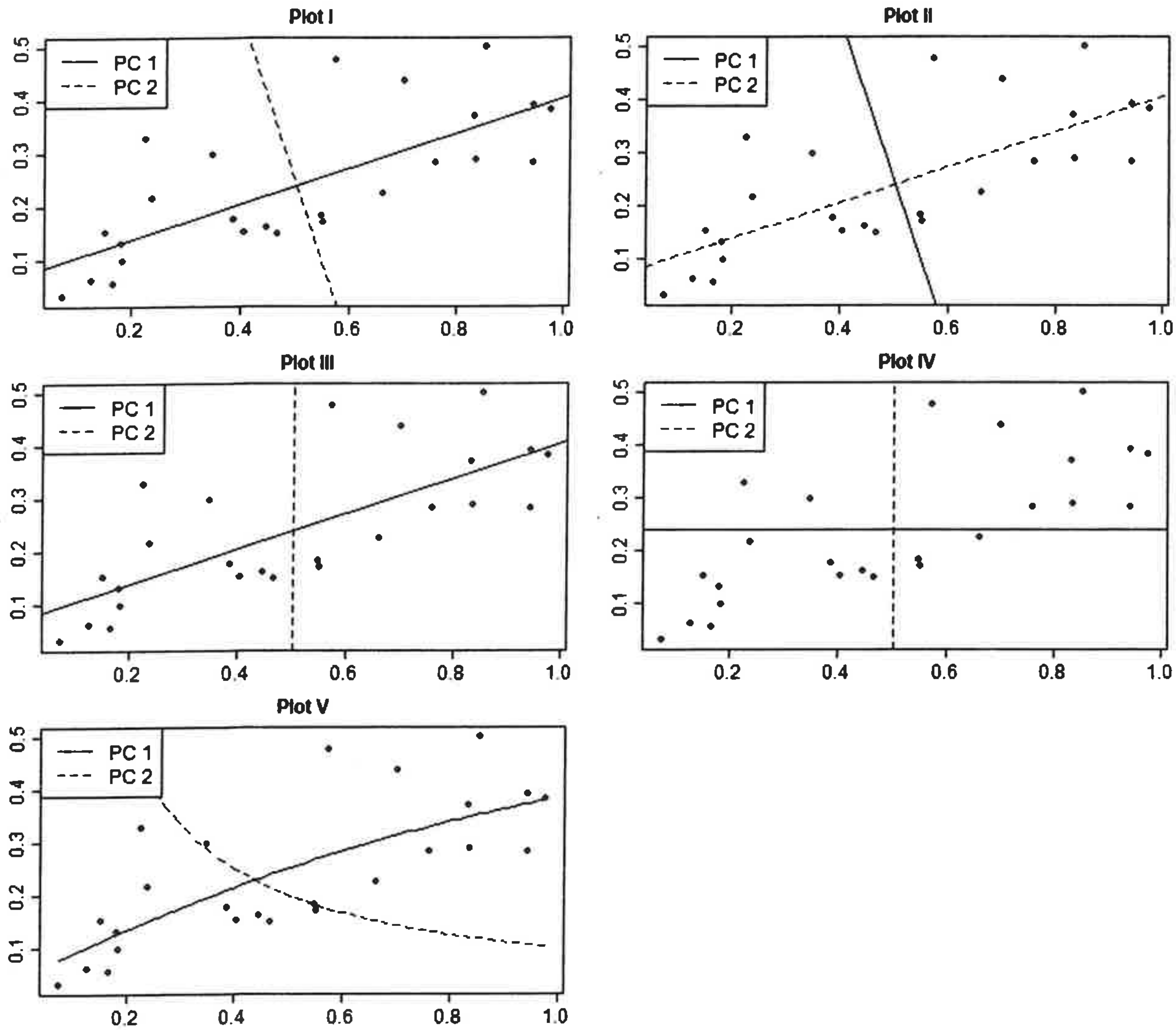
Determine which of the following statistics X and Y represent.

A. X = Squared Bias, Y = Training MSE
B. X = Test MSE, Y = Training MSE
C. X = Test MSE, Y = Variance
D. X = Training MSE, Y = Variance
E. X = Variance, Y = Test MSE

CONTINUED ON NEXT PAGE

38.

You are given a series of plots of a single data set containing two variables:



Determine which of above plots accurately represent the 1st and 2nd principal components (PC1 and PC2, respectively) of this dataset.

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

CONTINUED ON NEXT PAGE

39.

Two actuaries were given a dataset and asked to build a model to predict claim frequency using any of 5 independent predictors {1, 2, 3, 4, 5} as well as an intercept {I}.

- Actuary A chooses their model using Best Subset Selection
- Actuary B chooses their model using Forward Stepwise Regression
- When evaluating the models they both used R-squared to compare models with the same number of parameters, and AIC to compare models with different numbers of parameters.

Below are statistics for all candidate models:

Model	# of Non Intercept Parameters	Parameters	R^2	Log-likelihood
1	0	I	0	0.05
2	1	I, 1	0.56	1.3
3	1	I, 2	0.57	1.4
4	1	I, 3	0.55	1.2
5	1	I, 4	0.52	1.15
6	1	I, 5	0.51	1.1
7	2	I, 1, 2	0.61	2.5
8	2	I, 1, 3	0.64	2.75
9	2	I, 1, 4	0.63	2.6
10	2	I, 1, 5	0.69	3
11	2	I, 2, 3	0.61	2.5
12	2	I, 2, 4	0.62	2.55
13	2	I, 2, 5	0.68	2.9
14	2	I, 3, 4	0.66	2.8
15	2	I, 3, 5	0.64	2.75
16	2	I, 4, 5	0.6	2.45

Model	# of Non Intercept Parameters	Parameters	R^2	Log-likelihood
17	3	I, 1, 2, 3	0.73	3.35
18	3	I, 1, 2, 4	0.71	3.25
19	3	I, 1, 2, 5	0.72	3.3
20	3	I, 1, 3, 4	0.75	3.5
21	3	I, 1, 3, 5	0.76	3.6
22	3	I, 1, 4, 5	0.79	3.9
23	3	I, 2, 3, 4	0.78	3.7
24	3	I, 2, 3, 5	0.74	3.4
25	3	I, 2, 4, 5	0.75	3.45
26	3	I, 3, 4, 5	0.73	3.35
27	4	I, 1, 2, 3, 4	0.88	4.2
28	4	I, 1, 2, 3, 5	0.8	3.95
29	4	I, 1, 2, 4, 5	0.87	4.1
30	4	I, 1, 3, 4, 5	0.83	4
31	4	I, 2, 3, 4, 5	0.85	4.05
32	5	I, 1, 2, 3, 4, 5	0.9	4.25

- AIC_j is the AIC of the model chosen by Actuary j

Calculate the absolute value of the difference between AIC_A and AIC_B

A. Less than 0.15
B. At least 0.15, but less than 0.30
C. At least 0.30, but less than 0.45
D. At least 0.45, but less than 0.60
E. At least 0.60

CONTINUED ON NEXT PAGE

40.

A model is created with the following form:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - \xi)_+^3$$

where:

$$(x - \xi)_+^3 = (x - \xi)^3 \quad \text{if } x > \xi, \quad \text{and 0 otherwise}$$

You are given the following statements:

I. $f(x)$ is continuous at ξ

II. $f'(x)$ is continuous at ξ

III. $f''(x)$ is continuous at ξ

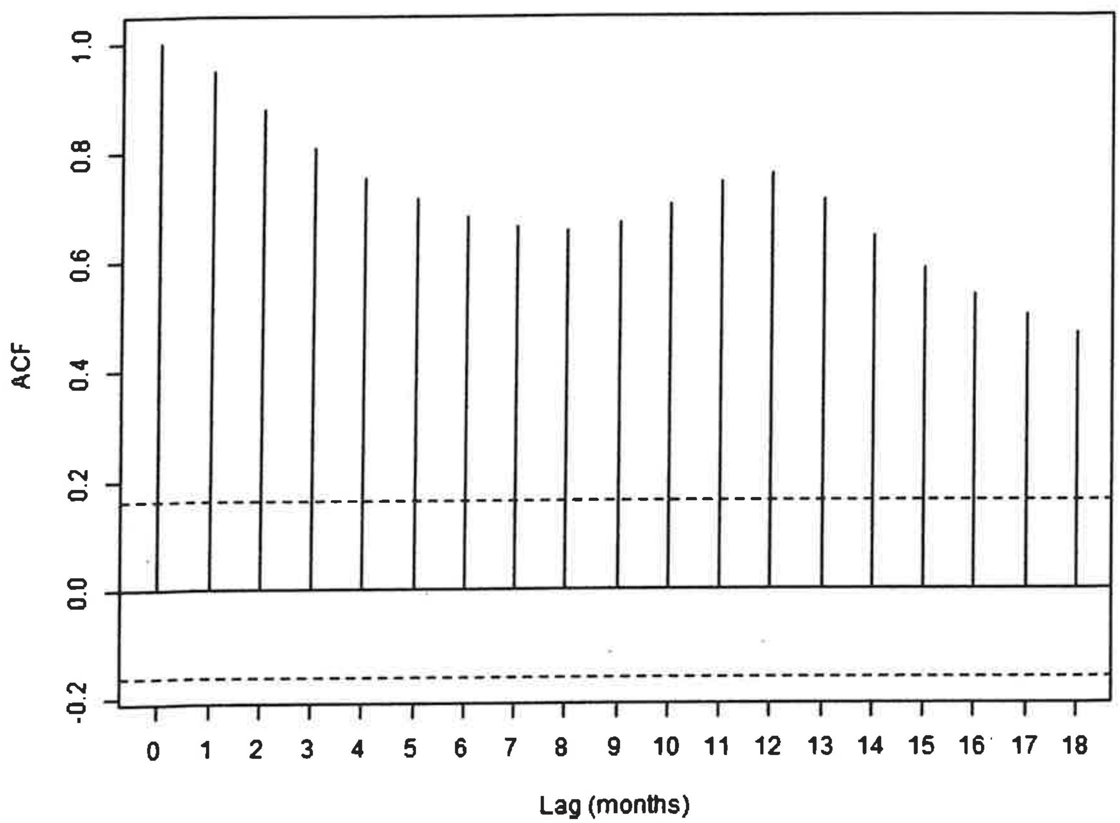
Determine which of the above statements are true.

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

41.

An actuary produces the following correlogram for vehicle accident severities over a 10-year period:



You are also given the following three statements about this time series:

I. There is a positive trend in accident severity
II. There is a negative trend in accident severity
III. The accident severity data shows a seasonal pattern

Determine which of the above statements can be conclude from the above graph.

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

42.

A time series is modeled using the function below:

$$x_t = \alpha_0 + \alpha_1 * t + z_t$$

- z_t is a white noise series
- $z_1 = -9.6$
- $z_3 = -4.7$
- The first order difference at time t = 2 is $\nabla x_2 = 26.2$
- The first order difference at time t = 3 is $\nabla x_3 = -1.3$
- $x_3 = 45.3$

Calculate the forecast value of x_7.

A. Less than 47
B. At least 47, but less than 67
C. At least 67, but less than 87
D. At least 87, but less than 107
E. At least 107

CONTINUED ON NEXT PAGE

43.

You are given the following time series model:

$$x_t = \frac{2}{3}x_{t-1} + \frac{1}{3}x_{t-2} + w_t$$

where $\{w_t\}$ is a white noise series.

Determine whether this time series is stationary and/or invertible.

A. Non-stationary, invertible
B. Non-stationary, not invertible
C. Stationary, invertible
D. Stationary, not invertible
E. The correct answer cannot be determined from the information given

CONTINUED ON NEXT PAGE

44.

You are given the following statements about time series and generalized least squares regression (GLS):

I. When there is positive serial correlation in a time series, the standard errors of the estimated regression parameters are likely to be over-estimated.
II. GLS is an improvement over ordinary least squares regression for serially correlated time series because GLS is based on maximizing the likelihood given the white noise in the data.
III. GLS can be used to provide better estimates of standard errors of the regression parameters to account for autocorrelation in the residual series.

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

45.

You are given the following quarterly rainfall totals over a two-year span:

Quarter	Rainfall
2016 q1	25
2016 q2	19
2016 q3	10
2016 q4	32
2017 q1	26
2017 q2	38
2017 q3	22
2017 q4	20

Calculate the sample lag 4 autocorrelation.

A. Less than 0.0
B. At least 0.0, but less than 0.3
C. At least 0.3, but less than 0.6
D. At least 0.6, but less than 0.9
E. At least 0.9

END OF EXAMINATION

Fall 2018 MAS-I Final Answer Key

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