Solutions to the
Fall 2019
CAS Exam MAS-1

(Incorporating the Final CAS Answer Key)

There were 45 questions in total, of equal value, on this 4 hour exam.
There was a 15 minute reading period in addition to the 4 hours.

prepared by
Howard C. Mahler, FCAS

Howard Mahler
hmahler@mac.com
www.howardmahler.com/Teaching

Where each question would go in my study guides:[1]

1. Stochastic Models, Section 2
2. Stochastic Models, Section 5
3. Stochastic Models, Section 10
4. Statistics, Section 31
5. Reliability Theory, Section 3
6. Reliability Theory, Section 4
7. Reliability Theory, Section 1
8. Stochastic Models, Section 18
9. Stochastic Models, Section 22
10. Stochastic Models, Section 26

11. Life Contingencies, Section 6
12. Life Contingencies, Section 2
13. Simulation, Section 4
14. Loss and Frequency Distributions, Section 23
15. Statistics, Section 38

16. Statistics, Section 2
17. Loss and Frequency Distributions, Section 38
18. Statistics, Section 6
19. Statistics, Section 23
20. Statistics, Section 15

21. Statistics, Section 21
22. Loss and Frequency Distributions, Section 20
23. Loss and Frequency Distributions, Section 29
24. Statistics, Section 31
25. Generalized Linear Models, Section 15

26. Generalized Linear Models, Section 5
27. Generalized Linear Models, Section 5
28. Statistical Learning, Section 7
29. Statistical Learning, Section 2
30. Regression, Section 16

[1] Some questions rely on ideas in more than one section of a study guide or even on ideas in more than one study guide. In those cases, I have chosen the best place to put a question in my opinion.

31. Regression, Section 15
32. Generalized Linear Models, Section 10
33. Generalized Linear Models, Section 8
34. Statistical Learning, Section 2
35. Regression, Section 3

36. Statistical Learning, Section 6
37. Statistical Learning, Section 5
38. Statistical Learning, Section 12
39. Statistical Learning, Section 4
40. Statistical Learning, Section 4

41. Time Series, Section 4
42. Time Series, Section 8
43. Time Series, Section 11
44. Time Series, Section 3
45. Time Series, Section 8

Out of a total of 45, the number of questions by my study guides:

	Number	Percent
Stochastic Models	6	13.3%
Loss & Freq. Dists.	4	8.9%
Statistics	8	17.8%
Regression	3	6.7%
GLMs	5	11.1%
Statistical Learning	8	17.8%
Life Contingencies	2	4.4%
Reliability	3	6.7%
Time Series	5	11.1%
Simulation	1	2.2%
Total	45	100.0%

1. Losses follow a memoryless distribution with mean 1,000.
Each loss is insured and subject to a deductible of 500.
Calculate the average insurance payment made on losses that exceed the deductible.
A. Less than 500
B. At least 500, but less than 700
C. At least 700, but less than 900
D. At least 900, but less than 1100
E. At least 1100

1. D. Since the distribution is memoryless, the average payment per payment is the mean of **1000**.
Comment: "Calculate the average insurance payment made on losses that exceed the deductible."
This language indicates to me the average payment per large loss, in other words per non-zero payment.

2. You are given the following information about waiting times at a subway station:
- Subway trains arrive at a Poisson rate of 20 per hour
- 30% of the trains are Express and 70% are Local
- The arrival times of each train are independent
- An Express train gets you to work in 18 minutes, and a Local train gets you there in 30 minutes
- You always take the first train to arrive and you get to the office in X_1 minutes from the time you arrive at the subway station
- Your coworker always takes the first Express train to arrive and he gets to the office in X_2 minutes from the time he arrives at the subway station

Calculate the expected value of $X_1 - X_2$.

A. Less than -2.0
B. At least -2.0, but less than -1.0
C. At least -1.0, but less than 0.0
D. At least 0.0, but less than 1.0
E. At least 1.0

2. E. Your average wait is: 1/20 hours = 3 minutes.
Your average ride is: (0.7)(30) + (0.3)(18) = 26.4 minutes.
X_1 = 3 + 26.4 = 29.4 minutes.
Express trains are Poisson with mean: (0.3)(20) = 6 per hour.
Your coworker's average wait is: 1/6 hours = 10 minutes.
X_2 = 10 + 18 = 28 minutes.
$X_1 - X_2$ = 29.4 - 28 = **1.4** minutes.
Comment: Similar to 3, 11/02, Q.20.

3. You are given the following information about an online retailer:

- Orders are placed on the website according to a homogeneous Poisson process with mean 50 per hour
- The number of items purchased in each order is independent and has the following distribution:

Number of Items	Probability
1	0.50
2	0.40
3	0.08
4	0.02

Calculate the variance of the total number of items purchased in a four-hour period.

A. Less than 100
B. At least 100, but less than 300
C. At least 300, but less than 500
D. At least 500, but less than 700
E. At least 700

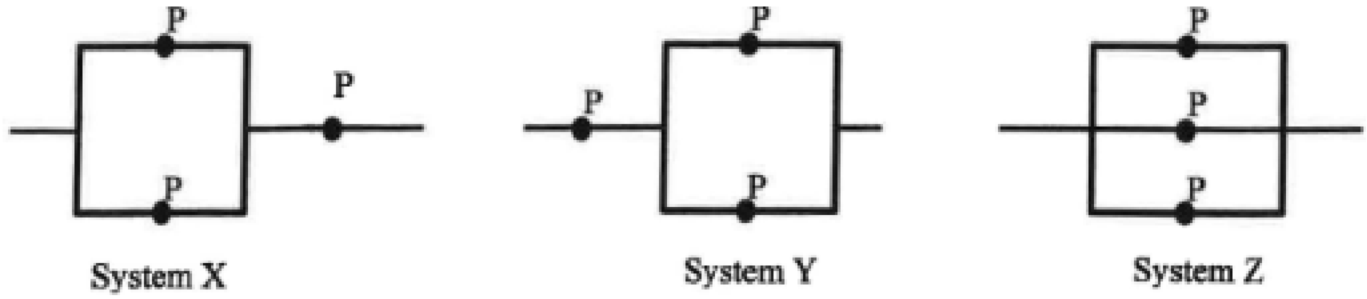
3. D. The second moment of severity is: $(0.5)(1^2) + (0.4)(2^2) + (0.08)(3^2) + (0.02)(4^2) = 3.14$.
Variance of the compound Poisson: (4)(50)(3.14) = **628**.

4. A building is powered by three generators with independent lifetimes, each following an exponential distribution with mean of one year. All three generators are started at the same time. Let T be the time in years between the fIrst and last generator failure.
Calculate E[T].
A. Less than 0.8
B. At least 0.8, but less than 1.2
C. At least 1.2, but less than 1.6
D. At least 1.6, but less than 2.0
E. At least 2.0

4. C. E[minimum] = (1)(1/3). E[maximum] = (1)(1/3 + 1/2 + 1).
T = E[maximum - minimum] = 1/2 + 1 = **1.5**.
Alternately, after the first generator fails, due to the memoryless property we have two remaining generators each of which follows an Exponential distribution with mean of one year.
The expected value of the maximum of these two Exponentials is: (1)(1/2 + 1) = **1.5**.
Comment: The r^{th} order statistic of an Exponential Distribution with mean θ is:

$$E[X_{(r)}] = \left\{\frac{1}{N} + \frac{1}{N-1} + ... + \frac{1}{N+1-r}\right\} \theta.$$

5. You are given the following bridge systems:



- All three systems X, Y, and Z are built with independent and identical components, P
- The reliability of System i is denoted as r_i for i = X, Y, Z

Determine which of the following best describes the reliabilities of these systems.

A. $r_X = r_Y > r_Z$

B. $r_X < r_Y < r_Z$

C. $r_X = r_Y = r_Z$

D. $r_X = r_Y < r_Z$

E. $r_X > r_Y > r_Z$

5. D. Systems X and Y have the same failure rate.
System Z has a lower failure rate, since this parallel system functions if any of its three components functions. $\mathbf{r_X = r_Y < r_Z}$.

<u>Comment</u>: Let p be the probability that component P functions.
In the trivial cases p = 0 or p = 1, all three systems have the same reliability.

$r_X = r_Y = p \{1 - (1-p)^2\}$. $r_Z = 1 - (1-p)^3$.

Let q = 1 - p. Since $0 < p < 1$, we have $0 < q < 1$.

$r_X = r_Y = (1 - q)(1 - q^2) = 1 - q - q^2 + q^3$. $r_Z = 1 - q^3$.

$r_Z - r_X = q + q^2 - 2q^3$. However, for $0 < q < 1$, $q > q^3$, and $q^2 > q^3$.

Thus $r_Z - r_X > 0$.

6. You are given a system which consists of the following minimal cut sets:

{1}, {2,3}, {4}, {5,6}

The system is comprised of independent and identically distributed components, each with reliability 0.9.

Calculate the lower bound of the reliability of the system by using the fIrst two inclusion-exclusion bounds from the method of inclusion and exclusion.

A. Less than 0.75

B. At least 0.75, but less than 0.77

C. At least 0.77, but less than 0.79

D. At least 0.79, but less than 0.81

E. At least 0.81

6. C. The chances that all components have failed of each minimal cut set are:

0.1, 0.1^2, 0.1, and 0.1^2.

$1 - r(p) \le 0.1 + 0.1^2 + 0.1 + 0.1^2$. $r(p) \ge 1 - 0.22 = 0.78$.

The chances that all components have failed of each pair of minimal cut sets are:

0.1^3, 0.1^2, 0.1^3, 0.1^3, 0.1^4, and 0.1^3.

$r(p) \le 0.78 + (0.1^3 + 0.1^2 + 0.1^3 + 0.1^3 + 0.1^4 + 0.1^3) = 0.7941$.

Thus if we stop after the first two bounds a lower bound for r(p) is **0.78**.

(There is no need to compute the 0.7941.)

Comment: In my opinion, a poorly constructed question.

Rather than testing whether you know the material, it tests whether under exam conditions you can get what the question writer had in mind. While it does not make much sense, the question did say "using the fIrst two inclusion-exclusion bounds."

Similar to poorly constructed CAS S, 5/17, Q.9.

Let C_i be the minimal cut sets for a system.

Let F_i = {all components of C_i are failed}.

$$1 - r(p) \le \sum_{i=1}^{n} P(F_i).$$

$$1 - r(p) \ge \sum_{i=1}^{n} P(F_i) - \sum_{j>i} P(F_i F_j).$$

$$1 - r(p) \le \sum_{i=1}^{n} P(F_i) - \sum_{j>i} P(F_i F_j) + \sum_{i=1}^{n}\sum_{j>i}\sum_{k>j} P(F_i F_j F_k).$$

The chances that all components have failed of each triplet of minimal cut sets are:

0.1^4, 0.1^5, 0.1^4, 0.1^5.

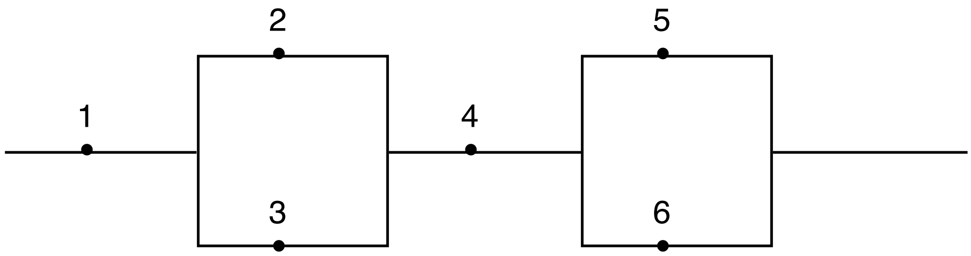
Thus, $r(p) \ge 0.78 - (0.1^4 + 0.1^5 + 0.1^4 + 0.1^5) = 0.79388$.

Thus a lower bound for r(p) if we stop after the first two lower bounds is 0.79388.

The chance that all components have failed of all four of the minimal cut sets is: 0.1^6.

Thus $r(p) = 0.79388 + 0.1^6 = 0.793881$.

The system diagram is:



$r(p) = (0.9)(1 - 0.1^2)(0.9)(1 - 0.1^2) = 0.793881$.

7. You are given the following information regarding a series system with two independent machines, X and Y:

- The hazard rate function, in years, for machine i is denoted by $r_i(t)$
- $r_X(t) = \ln(1.06)$, for $x > 0$
- $r_Y(t) = \frac{1}{20 - t}$, for $0 < y < 20$
- Both machines are currently three years old

Calculate the probability that the system fails when the machines are between five and nine years old.

A. Less than 0.305
B. At least 0.305, but less than 0.315
C. At least 0.315, but less than 0.325
D. At least 0.325, but less than 0:335
E. At least 0.335

7. D. $S(t) = \exp[-\int_0^t r(s)\, ds]$.

$S_X(t) = \exp[-t \ln[1.06]] = 1/1.06^t$.

$S_Y(t) = \exp[\ln(20-t) - \ln(20)] = (20 - t)/20$.

Prob[X survives until 5 I X survives until 3] = $S_X(5) / S_X(3) = (1/1.06^5) / (1/1.06^3) = 1/1.06^2$.

Prob[X survives until 9 I X survives until 3] = $S_X(9) / S_X(3) = 1/1.06^9) / (1/1.06^3) = 1/1.06^6$.

Prob[Y survives until 5 I X survives until 3] = $S_Y(5) / S_Y(3) = (15/20) / (17/20) = 15/17$.

Prob[X survives until 9 I X survives until 3] = $S_Y(9) / S_Y(3) = (11/20) / (17/20) = 11/17$.

The series system survives if both machines survive.

Prob[system survives until 5 I system survives until 3] = $(1/1.06^2)(15/17) = 0.7853$.

Prob[system survives until 9 I system survives until 3] = $(1/1.06^6)(11/17) = 0.4562$.

Probability that the system fails between five and nine: 0.7853 - 0.4562 = **0.3291**.

Comment: By the memoryless property, after time 3, X follows the same Exponential.

Prob[X survives until 9 I X survives until 3] = $1/1.06^2$.

Prob[X survives until 9 I X survives until 3] = $1/1.06^6$.

After time 3, Y is uniform for 17 more years.

Prob[Y survives until 5 I X survives until 3] = 15/17.

Prob[X survives until 9 I X survives until 3] = 11/17.

This exam question should have read: $r_X(t) = \ln(1.06)$, for $t > 0$, and $r_Y(t) = \frac{1}{20 - t}$, for $0 < t < 20$.

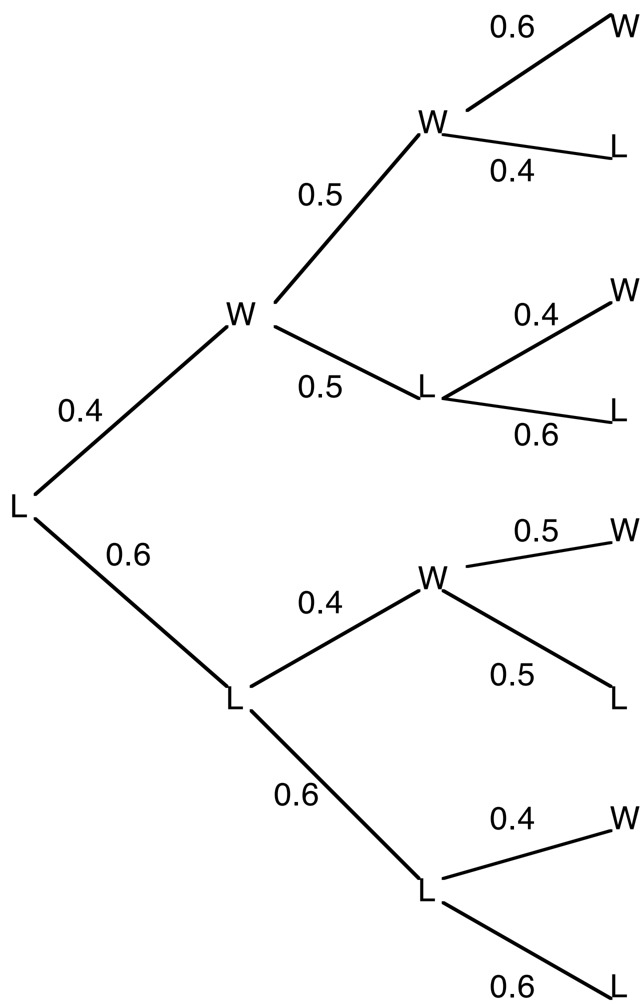
8. You are given:

- Mary plays a game repeatedly
- Each game ends with her either winning or losing
- Mary's chances of winning her next game depends on the outcome of the prior game
    - P[Winning after a win] = min(80%, P[Winning prior game] + 10%)
    - P[Winning after a loss] = 40%
- Mary just played her 10^{th} game and lost.

Calculate the probability that Mary will lose her 13^{th} game.

A. Less than 40%
B. At least 40%, but less than 45%
C. At least 45%, but less than 50%
D. At least 50%, but less than 55%
E. At least 55%

8. D. Not a simple Markov Chain.
If Mary wins game 11, then there is a 40% + 10% = 50% chance that she wins game 12.
If she also wins game 12, then there is a 50% + 10% = 60% chance that she wins game 13.
One can solve via a branch diagram:



The chance that Mary losses her 13th game is:
(0.4)(0.5)(0.4) + (0.4)(0.5)(0.6) + (0.6)(0.4)(0.5) + (0.6)(0.6)(0.6) = **0.536**.

Alternately, one could put this in the form of a Markov Chain.
State 1: Loss
State 2: Win preceded by a Loss
State 3: Win preceded by a Win preceded by a Loss
State 4: Win preceded by a Win preceded by a Win preceded by a Loss
State 5: Win preceded by a Win preceded by a Win preceded by a Win

Then the transition matrix is: $P = \begin{pmatrix} 0.6 & 0.4 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.4 & 0 & 0 & 0.6 & 0 \\ 0.3 & 0 & 0 & 0 & 0.7 \\ 0.2 & 0 & 0 & 0 & 0.8 \end{pmatrix}$.

(1, 0, 0, 0, 0) P = (0.6, 0.4, 0, 0, 0). (0.6, 0.4, 0, 0, 0) P = (0.56, 0.24, 0.20, 0, 0).
(0.56, 0.24, 0.20, 0, 0) P = (**0.536**, 0.224, 0.12, 0.12, 0).
<u>Comment</u>: Difficult and unusual.
(0.536, 0.224, 0.12, 0.12, 0) P = (0.5176, 0.2144, 0.112, 0.072, 0.084).
Thus the chance that Mary loses her 14^{th} game is: 0.5176.

9. You are given the following Markov chain transition probability matrix:

$$P = \begin{pmatrix} 0.0 & 0.3 & 0.3 & 0.4 & 0.0 \\ 0.4 & 0.0 & 0.0 & 0.6 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.5 & 0.5 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

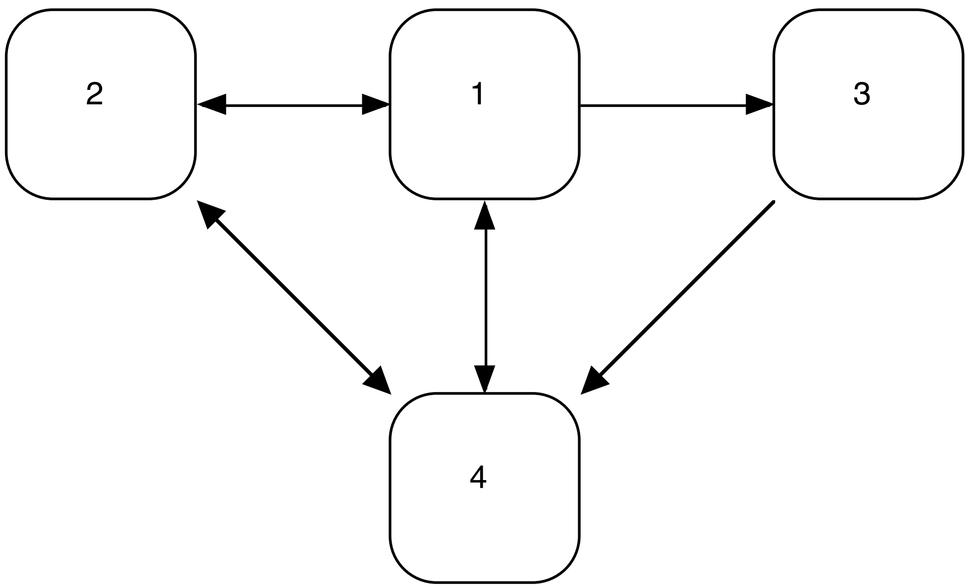
Determine the number of recurrent states in this Markov chain.

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

9. E. Label the states 1 to 5.

State 5 is an absorbing state in its own class; State 5 is recurrent.

For the other four states we have the following diagram:



These four states all communicate with each other and form a class; they are all recurrent.

All **5** states are recurrent.

<u>Comment</u>: A recurrent state is one such that if we are in that state there is zero chance we will never return to that state. A transient state is one such that if we are in that state, there is a chance we will never return to that state. All the states in a class are either recurrent or transient.

10. Ann and Beatrice are playing a game with multiple rounds:
- At the end of each round, the loser pays 1 coin to the winner
- Each player begins with 5 coins and plays until one of them has no coins remaining
- Beatrice is more experienced at the game, and her probability of winning each round is 0.75
- Each round is independent

Calculate the probability that Beatrice has no coins remaining at the end of the game.
A. Less than 0.01
B. At least 0.01, but less than 0.03
C. At least 0.03, but less than 0.05
D. At least 0.05, but less than 0.07
E. At least 0.07

10. A. Assuming the Gambler starts with i out of a total of N units of money and has a $p \neq 1/2$ chance of winning each game, then if $\beta = (1-p)/p$ = chance of failure / chance of success,

the chance of the Gambler ending up with all the money is: $(1 - \beta^i)/(1 - \beta^N)$.

Treating Beatrice as the Gambler: $\beta = 0.25/0.75 = 1/3$.

Chance that Beatrice ends up with all of the money = $\frac{1 - 1/3^5}{1 - 1/3^{10}}$ = 0.9959.

Probability that Beatrice has no coins remaining = 1 - 0.9959 = **0.0041**.
Alternately, treating Ann as the Gambler: $\beta = 0.75/0.25 = 3$.

Chance that Ann ends up with all of the money = $\frac{1 - 3^5}{1 - 3^{10}}$ = **0.0041**.

11. A company with extensive experience is using the Illustrative Life Table (ILT) to price life insurance, but is considering switching to a Modified ILT.
The Modified ILT is identical to the original ILT for Ages beyond 25, but has adjustments for Ages 25 and prior.
The following is an excerpt from the adjusted section of the Modified ILT:

Age x	l_x	$1000q_x$	$\ddot{a}_x$	$1000A_x$
20	9,617,802	1.34	16.5112	65.40
21	9,604,896	1.37	16.4640	68.08
22	9,591,695	1.10	16.4144	70.89
23	9,581,169	1.14	16.3572	74.12
24	9,570,288	0.86	16.2971	77.52
25	9,562,017	0.91	16.2290	81.38

This company only sells whole life policies of 10,000.
Someone Age 23 is considering purchasing a whole life policy.
Calculate the change in annual level benefit premium the company charges this customer by using the Modified ILT instead of the original ILT.
A. Less than -0.25
B. At least -0.25, but less than 0.00
C. At least 0.00, but less than 0.25
D. At least 0.25, but less than 0.50
E. At least 0.50

11. A. The level annual premium is: $10{,}000\ A_{23} / \ddot{a}_{23} = (10)(1000A_{23}) / \ddot{a}_{23}$.
Looking in the Illustrative Life Table: (10)(74.62) / 16.3484 = 45.64.
Looking in the Modified Illustrative Life Table: (10)(74.12) / 16.3572 = 45.31.
The change is: 45.31 - 45.64 = **-0.31**.
<u>Comment</u>: We do not make use of most of the information in the given Modified ILT.

12. An insurance company is using the Illustrative Life Table to price a block of life insurance policies covering 5,000 people aged 50.
Calculate the lower bound of a 90% confidence interval for the number of deaths in this block during the next 15 years, using a normal approximation.
A. Less than 745
B. At least 745, but less than 755
C. At least 755, but less than 765
D. At least 765, but less than 775
E. At least 775

12. B. l_{50} = 8,950,901. l_{65} = 7,533,964.
Thus the number of deaths is Binomial, with m = 5000 and q = 1 - 7,533,964/8,950,901 = 0.1583.
The mean is: (5000)(0.1583) = 791.5. The variance is: (5000)(0.1583)(1 - 0.1583) = 666.2.
Thus an approximate 90% confidence interval is: $791.5 \pm 1.645\sqrt{666.2}$ = **749** to 834.
Comment: Similar to Exercise 2.6 in "Life Contingencies Study Note for CAS Exam S".

13. An actuary is using the inversion method to simulate the waiting time until the 5^{th} event of a Poisson process with a rate $\lambda = 1$.
Five random draws from U(0, 1) are provided below:
0.2, 0.7, 0.8, 0.3, 0.5
Calculate the simulated waiting time until the 5^{th} event.
A. Less than 2.5
B. At least 2.5, but less than 3.5
C. At least 3.5, but less than 4.5
D. At least 4.5, but less than 5.5
E. At least 5.5

13. C. Each interevent time is Exponential with mean $1/\lambda = 1$.
Set $u = F(x) = 1 - e^{-x}$. $\Rightarrow$ $x = -\ln(1 - u)$.
The waiting times are: $-\ln(1 - 0.2) = 0.223$, 1.204, 1.609, 0.357, 0.693.
Their sum is: **4.086**.
Alternately, the waiting time until the 5th event is:
$-\ln[(1-0.2)(1-0.7)(1-0.8)(1-0.3)(1-0.5)]$ = **4.086**.
Comment: Given these random numbers, in this case one would get the same final answer instead setting $u = S(x)$ and thus $x = -\ln(u)$.
The waiting time until the 5^{th} event is a Gamma Distribution with $\alpha = 5$ and $\theta = 1/\lambda = 1$; this is how one would simulate such a Gamma Distribution.

14. You are given the following information:

- X is a random variable from a single-parameter Pareto distribution with $\alpha = 5$ and unknown θ
- $\bar{x}$ is the sample mean of n independent observations from this distribution
- $c\bar{x}$ is an unbiased estimator of θ

Calculate c.

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

14. A. The mean of a Single Parameter Pareto Distribution is: $\theta\ \alpha / (\alpha - 1),\ \alpha > 1$.

Thus $E[\bar{x}] = \theta\ 5/4$. $\Rightarrow$ $E[(4/5)\bar{x}] = \theta$. $\Rightarrow$ We want c = 4/5 = **0.8**.

Comment: For the Single Parameter Pareto, usually theta is known and we estimate alpha.

15. An actuary obtained the following random sample from a normal distribution with mean μ and variance σ^2:

1.16, -6.78, 4.04, 7.68, 0.85

Calculate the minimum variance unbiased estimate of σ^2.

A. Less than 15
B. At least 15, but less than 20
C. At least 20, but less than 25
D. At least 25, but less than 30
E. At least 30

15. D. For the Normal Distribution, maximum likelihood is the same as the method of moments; the maximum likelihood estimator of σ^2 is: $\frac{\sum (x_i - \bar{X})^2}{n}$.

Thus this is an asymptotically unbiased function of the sufficient statistics.
However, it is biased. Multiplying by n/(n-1) one gets the sample variance which is unbiased.
Therefore, the sample variance is unbiased estimator which is function of the sufficient statistics.
The Normal Distribution is a (two parameter) exponential family.

Therefore, the sample variance, $S^2 = \frac{\sum (x_i - \bar{X})^2}{n - 1}$,

is a Minimum Variance Unbiased Estimator of σ^2.

The sample mean is 1.39.
The sample variance is:

$$\frac{(1.39 - 1.16)^2 + (6.78 - 1.16)^2 + (4.04 - 1.16)^2 + (7.68 - 1.16)^2 + (0.85 - 1.16)^2}{5 - 1} = \mathbf{28.42}.$$

<u>Comment</u>: Somewhat similar to MAS-1, 11/18, Q.17; however, there the mean is known while here the mean is unknown.
One can use the STAT function of the electronic calculator to determine the sample variance.
The sample mean is a Minimum Variance Unbiased Estimator of μ.

16. Suppose that $X_1, ..., X_{10}$ is a random sample from a normal distribution with:

$$\sum_{i=1}^{10} X_i = 100 \text{ and } \sum_{i=1}^{10} X_i^2 = 2000$$

The parameters of this distribution are estimated using the method of moments with raw moments only.
Calculate the estimated variance of this distribution.
A. Less than 120
B. At least 120, but less than 140
C. At least 140, but less than 160
D. At least 160, but less than 180
E. At least 180

16. A. We set the theoretical mean equal to: $\sum_{i=1}^{10} X_i$ / 10 = 100/10 = 10.

We set the theoretical second moment equal to: $\sum_{i=1}^{10} X_i^2$ / 10 = 2000/10 = 200.

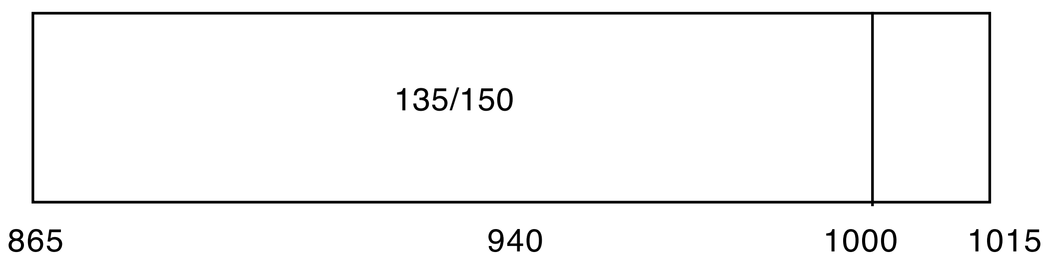
Thus the variance of the fitted distribution is: $200 - 10^2 =$ **100**.
Comment: The answer does not depend on it being a Normal Distribution; the answer would be the same for any two parameter distribution with a finite variance, for example a Gamma Distribution.
For the fitted Normal: $\hat{\mu} = 10$, and $\hat{\sigma} = \sqrt{100} = 10$.

17. You are given the following information:
- An insurer has observed the following nine losses:
  500, 600, 750, 880, 940, 1000, 1050, 1060, 1400
- A uniform kernel function with bandwidth 75 is used to estimate the distribution of loss sizes
- $\hat{F}(Y)$ is the kernel-smoothed estimate of the cumulative distribution function

Calculate $\hat{F}(1000)$.

A. Less than 0.55
B. At least 0.55, but less than 0.60
C. At least 0.60, but less than 0.65
D. At least 0.65, but less than 0.70
E. At least 0.70

17. C. The uniform kernels centered at the first four losses are completely to the left of 1000, and thus each contributes its whole area to $\hat{F}(1000)$. The uniform kernel centered at 1400 is completely to the right of 1000, and thus contributes nothing to $\hat{F}(1000)$.
The uniform kernel centered at 940 extends from 940 - 75 = 865 to 940 + 75 = 1015.



It contributes: (1000 - 865) / 150 = 135/150 of its area to $\hat{F}(1000)$.

The uniform kernel centered at 1000 contributes 1/2 of its area to $\hat{F}(1000)$.
The uniform kernel centered at 1050 extends from 975 to 1125.

It contributes: (1000 - 975) / 150 = 25/150 of its area to $\hat{F}(1000)$.
The uniform kernel centered at 1060 extends from 985 to 1135.

It contributes: (1000 - 985) / 150 = 15/150 of its area to $\hat{F}(1000)$.

$\hat{F}(1000)$ = (1 + 1 + 1 + 1 + 135/150 + 1/2 + 25/150 + 15/150 + 0) / 9 = **0.6296**.
Comment: Similar to 4, 11/04, Q.20.

18. You are given the following information about the distribution of losses:

- Losses follow an exponential distribution with mean θ.
- Insurance payments for each loss are subject to a deductible of 500 and a maximum payment of 30,000:

  Insurance payment = min[30,000, max(0, loss - 500)]
- No insurance payments are made for losses less than 500
- A random sample of five insurance payments are drawn:

  1,000 4,900 7,000 19,500 30,000

Calculate the maximum likelihood estimate of θ.

A. Less than 12,500

B. At least 12,500, but less than 13,500

C. At least 13,500, but less than 14,500

D. At least 14,500, but less than 15,500

E. At least 15,500

18. E. The final payment of 30,000 has been censored from above.

For maximum likelihood applied to the Exponential Distribution (with ungrouped data):

$$\hat{\theta} = \frac{\text{sum of the payments}}{\text{number of uncensored values}} = \frac{1,000 + 4,900 + 7,000 + 19,500 + 30,000}{4} = \mathbf{15,600}.$$

Alternately, the likelihood is: f(1000) f(4900) f(7000) f(19,500) S(30,000) =

$e^{-1000/\theta}/\theta \ e^{-4900/\theta}/\theta \ e^{-7000/\theta}/\theta \ e^{-19,500/\theta}/\theta \ e^{-30,000/\theta} = e^{-62,400/\theta}/\theta^4$.

Thus the loglikelihood is: $-62,400/\theta - 4\ln(\theta)$.

Setting the partial derivative with respect to theta equal to zero:

$0 = 62,400/\theta^2 - 4/\theta$. $\Rightarrow \theta = 62,400/4 =$ **15,600**.

<u>Comment</u>: Similar to 4, 5/07, Q.1.

19. You are testing the following hypotheses about a random variable, X:

- H_0: X is uniformly distributed on [0,16]
- H_1: X is uniformly distributed on [8,16]

You used a single observation and the best critical region with power of 0.95 to evaluate this hypothesis.

Calculate the probability of a Type I error.

A. Less than 0.40

B. At least 0.40, but less than 0.50

C. At least 0.50, but less than 0.60

D. At least 0.60, but less than 0.70

E. At least 0.70

19. B. For example take as a critical region [c, 16] where c ≥ 8.

Then the power is: Prob[observation is in the critical region | H_1] = (16 - c)/8 = 2 - c/8.

Probability of a Type I error = Prob[observation is in the critical region | H_0] = (16 - c)/16 = 1 - c/16.

We want power = 0.95. ⇒ 0.95 = 2 - c/8. ⇒ c = 8.4.

Probability of a Type I error = 1 - 8.4/16 = **0.475**.

<u>Comment</u>: Somewhat similar to CAS3, 11/06, Q.6.

In fact any subset of [8, 16] with total length of 7.6 will be a critical region with probability of a Type I error equal to 0.475. These are each best critical regions with power 0.95, in other words with the smallest probability of a Type I error for the given power.

[7, 15.6] is a critical region with power of 0.95, but with a probability of a Type I error of:
(15.6 - 7)/16 = 0.5375 > 0.475.

20. An insurance company has classified claims into 5 categories based on their severity.
The null hypothesis H_0 assumes the numbers of claims for Categories 1, 2, 3, 4 and 5 appear in the following ratios:

12: 8: 6: 4: 1

In 2017, the insurance company recorded the numbers of claims as follows:

Category	# of Claims
1	1172
2	829
3	605
4	347
5	102

Calculate the Chi-square goodness-of-fit statistic.

A. Less than 5.5
B. At least 5.5, but less than 6.5
C. At least 6.5, but less than 7.5
D. At least 7.5, but less than 8.5
E. At least 8.5

20. D. The total observed is 3055. The total of the proportions is 31.
For example: (3055)(12/31) = 1182.58. $(1172 - 1182.58)^2$ / 1182.58 = 0.095.

Number of Claims	Observed # of Claims	Assumed Distribution	Expected # Insureds	Chi-Square = (observed - expected)^2 / expected
1	1172	0.387097	1,182.58	0.095
2	829	0.258065	788.39	2.092
3	605	0.193548	591.29	0.318
4	347	0.129032	394.19	5.650
5	102	0.032258	98.55	0.121
Sum	3,055	1.000000	3,055.00	**8.276**

Comment: There are 5 - 1 = 4 degrees of freedom. The 5% critical value is 9.49.
Since 8.276 < 9.49, at the 5% significance level we do not to reject the null hypothesis.

21. An insurer is estimating the impact of a loss mitigation program. They ran an experiment to evaluate the severities of five losses before and after the program was instituted:

	A	B	C	D	E
Original Severity	400	800	1200	2000	5000
New Severity	280	500	1235	1600	4800

A paired t-test with the following hypotheses was used to evaluate the effectiveness of this program:

- H_0: The program had no impact on losses
- H_1: The program was able to reduce losses

Calculate the smallest significance level at which one would reject the null hypothesis.

A. Less than 1.0%
B. At least 1.0%, but less than 2.5%
C. At least 2.5%, but less than 5.0%
D. At least 5.0%, but less than 10.0%
E. At least 10.0%

21. C. Due to the form of H_1, we perform a one-sided t-test.

We work with the differences between the original and new severities:
120, 300, -35, 400, 200.
These differences have a sample mean 197, and sample variance 27,895.

$$t = \frac{197}{\sqrt{27,895/5}} = 2.637, \text{ with } 5 - 1 = 4 \text{ degrees of freedom.}$$

2.132 < 2.637 < 2.776.
Thus for a one-sided test, the p-value is between 5% and 2.5%.

22. You are given the following information about a sample, $X_1 \dots X_n$:

- X_i's are all mutually independent
- $X_i \sim \text{Gamma}(\alpha_i, \theta)$, for i = 1, 2, ... n
- $\alpha_i = \frac{1}{n}$ for all i
- $Y = \sum_{i=1}^{n} X_i$

Calculate the probability that $Y > \theta$.

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

22. B. Y is Gamma(1/n + ... + 1/n, θ) = Gamma (1, θ) = Exponential(θ).

Prob[$Y > \theta$] = $\exp[-\theta/\theta] = e^{-1}$ = **0.368**.

Comment: The sum of independent Gamma variables each with the same theta, is another Gamma, with the new alpha being the sum of the individual alphas.

23. You are given the following information about losses covered by an insurance company:

- Individual losses follow a lognormal distribution with ($\mu = 8$, $\sigma = 1$)
- For insurance payments, an ordinary deductible of 1,000 per loss applies
- Losses below the deductible are not reported to the insurance company

Calculate the mean payment made by the insurance company.

A. Less than 3,900

B. At least 3,900, but less than 4,200

C. At least 4,200, but less than 4,500

D. At least 4,500, but less than 4,800

E. At least 4,800

23. D. $S(1000) = 1 - \Phi\left[\frac{\ln(1000) - 8}{1}\right] = 1 - \Phi[-1.09] = 0.8621.$

$E[X] = \exp[\mu + \sigma^2/2] = \exp[8 + 1^2/2] = \exp[8.5] = 4914.8.$

$E[X \wedge x] = \exp(\mu + \sigma^2/2)\,\Phi\left[\frac{\ln(x) - \mu - \sigma^2}{\sigma}\right] + x\left\{1 - \Phi\left[\frac{\ln(x) - \mu}{\sigma}\right]\right\}$

$E[X \wedge 1000] = 4914.8\,\Phi\left[\frac{\ln(1000) - 8 - 1^2}{1}\right] + 1000\left\{1 - \Phi\left[\frac{\ln(1000) - 8}{1}\right]\right\} =$

$4914.8\,\Phi[-2.09] + 1000\,(1 - \Phi[-1.09]) = (4914.8)(1 - 0.9817) + (1000)(0.8621) = 952.0.$

Average payment per payment $= \frac{E[X] - E[X \wedge 1000]}{S(1000)} = \frac{4914.8 - 952.0}{0.8621} = \mathbf{4597}.$

Comment: "Losses below the deductible are not reported to the insurance company."
Therefore, from the point of view of the insurer, the mean payment they make is per non-zero payment. The insurer does not know how many losses there are, and therefore would not know the mean payment per loss. Nevertheless, it would have been nice for the question to be clearer that they are looking for the average payment per non-zero payment.
If one has data truncated from below (or truncated and shifted from below), then one can still fit an assumed ground up distribution (for example a LogNormal) via maximum likelihood.

24. A student would like to estimate the upper bound of a uniform distribution, U(0, θ) using the method of maximum likelihood.

The true value of θ is 10.

N is the minimum sample size required such that the absolute value of the bias of the estimator is less than 0.1.

Calculate N.

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

24. E. $\hat{\theta}$ = Max[X_i]. $\Rightarrow$ E[$\hat{\theta}$] = θ N/(N+1). $\Rightarrow$ Bias = θ N/(N+1) - θ = -θ/(N+1) = -10/(N+1).

Set the absolute value of the bias equal to 0.1: 0.1 = 10/(N+1). $\Rightarrow$ N = 99.

For the absolute value of the bias to be less than 0.1, we need N at least equal to **100**.

25. A bank uses a logistic model to estimate the probability of clients defaulting on a loan, and it comes up with the following parameter estimates:

i	Variable	β_i
0	Intercept	-1.6790
1	Income (in 000's)	-0.0294
2	Student [Yes]	-0.3870
3	Number of credit cards	0.7710

The following four clients applied for loans from the bank:

Client	Income	Student	# of credit cards
1	25,000	Y	1
2	10,000	Y	3
3	20,000	N	0
4	75,000	N	3

The bank will reject any loan if the probability of default is greater than 10%.
Calculate the number of clients whose loan requests are rejected.

A. 0 B. 1 C. 2 D. 3 E. 4

25. D. Chance of default = $\exp[X\beta] / (1 + \exp[X\beta])$.

Client 1: $X\beta$ = -1.6790 + (25)(-0.0294) - 0.3870 + (1)(0.7710) = -2.030.

exp[-2.030] / (1 + exp[-2.030]) = 11.6% > 10%.

Client 2: $X\beta$ = -1.6790 + (10)(-0.0294) - 0.3870 + (3)(0.7710) = -0.047.

exp[-0.047] / (1 + exp[-0.047]) = 48.8% > 10%.

Client 3: $X\beta$ = -1.6790 + (20)(-0.0294) = -2.267.

exp[-2.267] / (1 + exp[-2.267]) = 9.39% < 10%.

Client 4: $X\beta$ = -1.6790 + (75)(-0.0294) + (3)(0.7710) = -1.571.

exp[-1.571] / (1 + exp[-1.571]) = 17.21% > 10%.

3 clients loan requests are rejected.

<u>Comment</u>: $0.1 = \exp[X\beta] / (1 + \exp[X\beta]) \Rightarrow X\beta = -\ln(9) = -2.197$.

Therefore, for $X\beta > -2.197$, the chance of default is greater than 10%.

26. A number of candidate models were fit using the following variables:
- An intercept term
- Variable A - a Yes/No indicator
- Variable B - a Yes/No indicator
- An interaction of Variables A and B

There are four observations, which were arranged into the following design matrix:

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

This data was fit using three different link functions:
I. Identity
II. Inverse
III. Log
The predicted values, given below, were the same under all three models:

$$\hat{Y} = \begin{pmatrix} 0.50 \\ 0.80 \\ 0.40 \\ 0.70 \end{pmatrix}$$

Determine for which of the above link functions the estimated interaction coefficient is non-zero.
A. Identity, Inverse and Log
B. Identity and Inverse only
C. Identity and Log only
D. Inverse and Log only
E. The answer is not given by (A), (B), (C) or (D)

26. D. For the Identity Link Function: $\beta_0 = 0.5$. $\beta_0 + \beta_A = 0.8$. $\Rightarrow \beta_A = 0.3$.

$\beta_0 + \beta_B = 0.4$. $\Rightarrow \beta_B = -0.1$. $\beta_0 + \beta_A + \beta_B + \beta_{AB} = 0.7$. $\Rightarrow \beta_{AB} = 0.7 - 0.5 - 0.3 - (-0.1) = 0$.

For the Inverse Link Function: $\beta_0 = 1/0.5 = 2$. $\beta_0 + \beta_A = 1/0.8 = 1.25$. $\Rightarrow \beta_A = -0.75$.

$\beta_0 + \beta_B = 1/0.4 = 2.5$. $\Rightarrow \beta_B = 0.5$. $\beta_0 + \beta_A + \beta_B + \beta_{AB} = 1/0.7$.

$\Rightarrow \beta_{AB} = 1/0.7 - 2 - (-0.75) - 0.5 = -0.321 \neq 0$.

For the Log Link Function: $\beta_0 = \ln(0.5)$. $\beta_0 + \beta_A = \ln(0.8)$. $\Rightarrow \beta_A = \ln(0.8) - \ln(0.5) = \ln(1.6)$.

$\beta_0 + \beta_B = \ln(0.4)$. $\Rightarrow \beta_B = \ln(0.4) - \ln(0.5) = \ln(0.8)$.

$\beta_0 + \beta_A + \beta_B + \beta_{AB} = \ln(0.7)$. $\Rightarrow \beta_{AB} = \ln(0.7) - \ln(0.5) - \ln(1.6) - \ln(0.8) = 0.0896 \neq 0$.

Thus the estimated interaction coefficient is non-zero for the **Inverse and Log only**.

Comment: It would be extremely unusual to get the same predicted values using three different link functions.

27. An actuary is asked to model a non-negative response variable and requires that the model form produces an unbiased estimate.
Determine which error structure and link function combination would be the best choice for the modeling request.
A. Poisson and Identity
B. Compound Poisson-Gamma and Log
C. Normal and Identity
D. Gamma and Log
E. Poisson and Log

27. E. Since the response variable is non-negative, we do not wish to use the Normal Distribution.
Using the canonical link function makes the estimate from the GLM unbiased.
The canonical link function for the Poisson is the Log.
Comment: See the CAS study note by Michael Larsen.
The canonical link function for the Normal is the Identity.
The canonical link function for the Gamma is the Inverse.
The Tweedie Distribution for $1 < p < 2$ is a Compound Poisson-Gamma distribution.

28. Determine which one of the following statements about Principal Component Regression (PCR) is FALSE.

A. When performing PCR it is recommended that the modeler standardize each predictor prior to generating the principal components.

B. PCR is useful for performing feature selection.

C. PCR assumes that the directions in which features show the most variation are the directions that are associated with the target.

D. PCR can reduce overfitting.

E. The fIrst principal component direction of the data is that along which the observations vary the most.

28. B. While PCR can reduce the number of dimensions, the principal components usually do not have an intuitive meaning, and therefore cannot be used for feature selection.

Comment: Each of the principal components used in the regression is a linear combination of all of the original features.

If the assumption in statement C is a reasonable enough approximation, then Principal Component Regression will give good results.

Unlike PCR, partial least squares (PLS) identifies linear combinations of the features that not only approximate these features well, but also that are related to the response.

29. An actuary has a dataset with four observations and wants to use Leave-One-Out Cross Validation (LOOCV) to determine which one of two competing models fits the data better.
The model preference will be based on minimizing the mean squared error.
The values of the dependent variable are:

$y = [y_1, y_2, y_3, y_4] = [1.55, 1.55, 1.60, 1.95]$

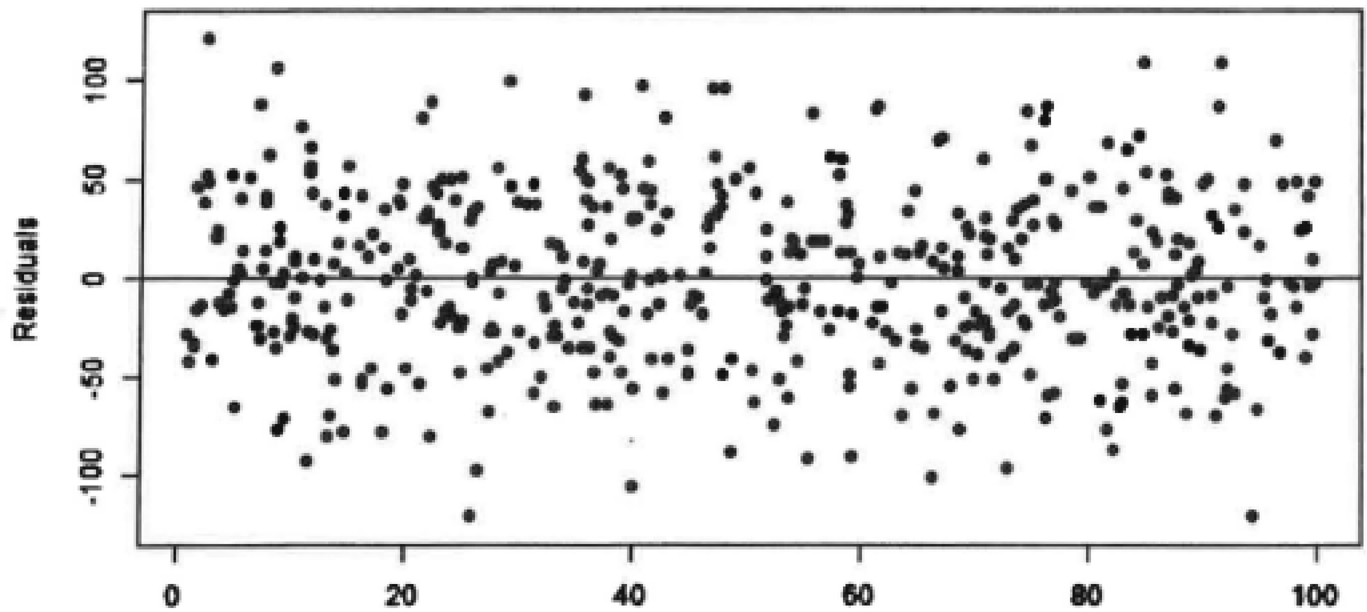
Corresponding fitted values under each model and training data subset are:

Training Obs. Used	Model 1				Model 2			
	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	$\hat{y}_4$	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	$\hat{y}_4$
1,2,3	1.50	1.60	1.20	1.80	1.60	1.70	1.60	Z
1,2,4	2.00	1.50	1.10	1.90	1.80	1.40	1.30	1.70
1,3,4	1.75	1.55	1.70	2.10	1.40	1.30	1.50	1.95
2,3,4	1.70	1.65	1.60	2.00	1.60	1.70	1.20	2.00

Calculate the maximum value of Z for which the actuary will prefer Model 2.
A. Less than 1.5
B. At least 1.5, but less than 1.8
C. At least 1.8, but less than 2.1
D. At least 2.1, but less than 2.4
E. At least 2.4

29. D. In each case, we compare the predicted and actual value for the y_i in the test set, in other words not in the training set. For example, for the first row, $y_4 = 1.95$ is the test set; for the first Model the contribution to the squared error is: $(1.80 - 1.95)^2$.
For Model 1, the test mean squared error is:
$(1/4) \{(1.80 - 1.95)^2 + (1.10 - 1.60)^2 + (1.55 - 1.55)^2 + (1.70 - 1.55)^2\} = 0.07375$.
For Model 2, the test mean squared error is:
$(1/4) \{(Z - 1.95)^2 + (1.30 - 1.60)^2 + (1.30 - 1.55)^2 + (1.60 - 1.55)^2\} = 0.03875 + (1/4) (Z - 1.95)^2$.
Setting the two mean squared errors equal:
$0.07375 = 0.03875 + (1/4) (Z - 1.95)^2. \Rightarrow (Z - 1.95)^2 = 0.14. \Rightarrow Z = 1.95 \pm 0.374$.
Thus the maximum possible value of Z is: 1.95 + 0.374 = **2.324**.
Comment: For values of Z more than 2.324 or less than 1.576, Model 2 would have a larger test mean squared error than Model 1.

30. An actuary has a data set with one predictor variable, X, and a response variable, Y.
She divides the data set randomly into training and testing sets. The training subset is used to fit an ordinary least squares regression. In order to evaluate the fit, she plots the residuals from the model against the independent variable, X:



Determine which of the following enhancements to the model would most likely improve the fit to the testing data set.
A. Linear Spline
B. Local Regression
C. Polynomial Regression
D. Step Function
E. There is no evidence that any of (A), (B), (C), (D) will improve the fit

30. E. I do not detect any pattern in the residuals.
Thus there is no evidence that any of the given choices would improve the fit.

31. In order to predict individual candidates' test scores a regression was performed using one independent variable, Hours of Study, plus an intercept.
Below is a partial table of data and model results:

		Hours of		Standardized
Candidate	Test Score	Study	Leverage	Residuals
1	2,041	538	0.6205	-1.3477
2	2,502	548	0.2018	-0.4171
3	2,920	528	0.6486	-1.1121
4	2,284	608	0.2807	1.1472

Calculate the number of observations above that are influential using Cook's Distance with a unity threshold.

A. 0 B. 1 C. 2 D. 3 E. 4

31. C. Cook's Distance = (i^{th} standardized residual)^2 $\frac{h_{ii}}{(k+1)\ (1 - h_{ii})}$,

where k is the number of fitted slope, in this case one for Hours of Study.

For example: $(-1.3477)^2 \frac{0.6205}{(1 + 1)\ (1 - 0.6205)} = 1.485$.

		Standardized	Cook's
Candidate	Leverage	Residuals	Distance
1	0.6205	-1.3477	1.485
2	0.2018	-0.4171	0.022
3	0.6486	-1.1121	1.141
4	0.2807	1.1472	0.257

The first and third observations have Cook's distances greater than 1 and are influential.
Two observations are influential.
Comment: One needs a combination of a large absolute value of the standardized residual and a large leverage (close to one), in order for a point to be influential.

32. You have three competing GLMs that each predict the number of claims under an insurance policy, and are evaluating the models using AIC and BIC. All models are trained on the same dataset of 300 observations. These models are summarized below:

Model	Likelihood	Number of Parameters
1	0.0456	4
2	0.0567	5
3	0.0575	6

The following are three statements about the fit of these models:
I. Model #1 is best based on BIC
II. Model #2 is best based on AIC
ill. Model #3 is best based on BIC
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)

32. A. AIC = (-2) (maximum loglikelihood) + (number of parameters)(2).
BIC = (-2) (maximum loglikelihood) + (number of parameters) ln(number of data points)
= (-2) ln(likelihood) + (number of parameters) ln(300).

Model	Likelihood	Number of Parameters	AIC	BIC
1	0.0456	4	**14.176**	**28.991**
2	0.0567	5	15.740	34.259
3	0.0575	6	17.712	39.935

Smallest AIC is best, which is Model 1.
Smallest BIC is best, which is Model 1.
Comment: Since Statements I and III cannot both be true, choice D is impossible.

33. You have a sample of five independent observations, x_1 ... x_5, each with exponential distribution: $f(x_i \mid \theta_i) = \frac{1}{\theta_i} \exp(-\frac{x_i}{\theta_i})$

- You are fitting this data to a model with $\theta_i = \theta$, for all i, using maximum likelihood estimation:

$$f(x_i \mid \theta) = \frac{1}{\theta} \exp(-\frac{x_i}{\theta})$$

- The five observed values are: 100, 100, 500, 800, 1000
- The deviance of the model, D, is equal to twice the difference between the log-likelihood of the saturated model and the log-likelihood of the fitted model.

Calculate D.

A. Less than 2

B. At least 2, but less than 4

C. At least 4, but less than 6

D. At least 6, but less than 8

E. At least 8

33. C. The maximum likelihood Exponential has $\theta = \bar{X} = 500$.

$\ln f(x_i) = -x_i/500 - \ln(500)$.

Loglikelihood is: $\sum \ln f(x_i) = -(2500/500) - 5\ln(500) = -36.07$.

The saturated model has as many parameters as the number of observations.
The saturated model has the largest possible likelihood, of models of a given form.
In this case, we want $\theta_i = x_i$.

(The maximum likelihood is when each θ_i is equal to the mean of its sample of size one.)

$\ln f(x_i) = -x_i/\theta_i - \ln(\theta_i) = -1 - \ln(x_i)$.

Loglikelihood is: $\sum \ln f(x_i) = -5 - \{\ln(100) + \ln(100) + \ln(500) + \ln(800) + \ln(1000)\} = -34.02$.

D = (2){-34.02 - (-36.07)} = **4.10**.

Alternately, not a Generalized Linear Model, yet one can use the formula for the Deviance for a Gamma error function: $D = 2\,\alpha \sum_{i=1}^{n} \{-\ln[y_i / \hat{y}_i] + (y_i - \hat{y}_i) / \hat{y}_i\}$.

We take $\alpha = 1$ since we have an Exponential. For the fitted Exponential $\hat{y}_i = 500 = \bar{X}$ for all i.

Thus here: $\sum_{i=1}^{n} (y_i - \hat{y}_i) / \hat{y}_i = 0$.

D = (2)(1) {-ln(100/500) - ln(100/500) - ln(500/500) - ln(800/500) - ln(1000/500)} = **4.11**.

34. You are given the following statements comparing k-fold cross validation (with k < n) and Leave-One-Out Cross Validation (LOOCV), used on a GLM with log link and gamma error.
I. k -fold validation has a computational advantage over LOOCV
II. k-fold validation has an advantage over LOOCV in bias reduction
III. k-fold validation has an advantage over LOOCV in variance reduction
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)

34. C. Statement I is true.
The test MSE estimated using these approaches tends to overestimate the test error rate for the model eventually fit on the entire original data set. Since LOOCV fits to all but one of the elements, it has less bias than k-fold validation. Thus Statement II is backwards.
The models fit in LOOCV are fit to virtually the same data; the data used differs by only two observations. Thus these models are highly correlated. Thus when we average the corresponding MSEs there is a high variance. The models fit in k-fold cross-validation are fit to similar data; however, the data used differs by twice the number of observations in each fold. Thus these models are less correlated; when we average the corresponding MSEs there is a lower variance than was the case for LOOCV. Statement III is true.
Comment: Since we have a log link and a gamma error, we do not have least squares linear or polynomial regression, for which LOOCV has a shortcut formula, and Statement I would not be true.

35. Two variables, X and Y, exhibit the following relationship:

$Y_i = 1.5X_i + 2 + \varepsilon_i$

where ε_i is a standard normal random variable, and each ε_i is mutually independent.
For some sample of data, an actuary uses ordinary least squares regression of the form:

$Y_i = \hat{\beta}_1 X_i + \hat{\beta}_0 + e_i$

to estimate the relationship. The following parameter estimates were formed:

Parameter	Estimate
$\hat{\beta}_0$	2.5
$\hat{\beta}_1$	1.3

Calculate the bias of the estimate.

A. Less than -0.1
B. At least -0.1, but less than 0.1
C. At least 0.1, but less than 0.3
D. At least 0.3
E. There is not enough information provided to calculate the bias.

35. B & E. (See Comment) $E[\hat{Y}_i] = E[1.3X_i + 2.5] = 1.3X_i + 2.5$.

Bias = $E[\hat{Y}_i] - E[Y_i] = (1.3X_i + 2.5) - (1.5X_i + 2) = -0.2X_i + 0.5$.

Thus there is not enough information provided to calculate the bias.
Comment: The CAS also allowed B, corresponding to zero bias.
Given the assumptions, least squares regression is an unbiased estimator; in other words a priori the bias is zero. However, given the particular parameter estimates as well as the true values of the parameters, this estimate is biased (for most values of X_i).
In my opinion, a poorly thought out question.
Usually ε_i is a normal random variable with mean 0; however, while all of the ε_i's are assumed to have the same variance, they need not have variance of 1 as stated in this exam question.

36. You are given the following three statements regarding shrinkage methods in linear regression:

I. As tuning parameter, λ , increases towards ∞, the penalty term has no effect and a ridge regression will result in the unconstrained estimates.

II. For a given dataset, the number of variables in a lasso regression model will always be greater than or equal to the number of variables in a ridge regression model.

III. The issue of selecting a tuning parameter for a ridge regression can be addressed with cross-validation.

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) .

36. C. Statement I describes what would happens as instead λ decreases towards zero.

Statement II is backwards.

Statement III is true.

Comment: Statement III is also true for the lasso.

Since using the lasso for large enough lambda some fitted coefficients are zero, the number of variables in a lasso regression model will always be less than or equal to the number of variables in a ridge regression model.

37. For a set of data with 40 observations, 2 predictors (X_1 and X_2), and one response (Y), the residual sum of squares has been calculated for several different estimates of a linear model with an intercept. Only integer values from 1 to 3 were considered for estimates of β_0 (the intercept), β_1 and β_2.

The grid below shows the residual sum of squares for every combination of the parameter estimates, after standardization:

		$\hat\beta_0 = 1$ $\hat\beta_2$			$\hat\beta_0 = 2$ $\hat\beta_2$			$\hat\beta_0 = 3$ $\hat\beta_2$		
		1	2	3	1	2	3	1	2	3
$\hat\beta_1$	1	3,924	1,977	1,250	3,949	1,822	1,174	3,784	1,671	1,107
	2	1,858	1,141	711	1,907	1,187	717	1,827	1,128	668
	3	1,386	822	369	1,363	711	349	1,294	700	344

Let $\hat\beta_i^R$ be the estimate of β_i using a ridge regression with budget parameter s = 5.

Assume the intercept is not subject to the budget parameters.

Calculate the value of $\hat\beta_0^R + \hat\beta_1^R + \hat\beta_2^R$.

A. Less than 6
B. 6
C. 7
D. 8
E. Greater than 8

37. B. The constraint for this ridge regression is: $\hat\beta_1^2 + \hat\beta_2^2 \le 5$.

We can have: $(\beta_1, \beta_2) = (1,1), (1,2)$, or $(2,1)$.

Given the constraint, the smallest residual sum of squares is for: $\hat\beta_0^R = 3$, $\hat\beta_1^R = 1$, and $\hat\beta_2^R = 2$.

Comment: Similar to MAS-1, 5/18, Q.36, which instead involves the lasso.

Given the constraint, the maximum value of $\hat\beta_0^R + \hat\beta_1^R + \hat\beta_2^R$ is 6.

The intercept is never subject to the budget parameters.

If instead we had the lasso, then the constraint with s = 5 would have been: $\hat\beta_1 + \hat\beta_2 \le 5$.

38. A modeler creates a cubic spline model with Property Claim Frequency as the response variable and Age of Building Construction as a continuous predictor.
The modeler puts knots in at Age = {10, 20, 30, 50}.
The modeler believes that she is overfitting on the ends of the distribution and decides to impose an additional constraint that the curve before the fIrst knot and after the last knot will be linear.
Calculate the number of degrees of freedom used by this new model.
A. 3 B. 4 C. 5 D. 6 E. 7

38. B & D. (See Comment)
There are 4 knots; K = 4. The constraint makes this a natural cubic spline.
A natural cubic spline has K degrees of freedom or **4**.
<u>Comment</u>: The CAS also allowed D, which corresponds to K + 2 = 6 degrees of freedom.
I do not know the reason for this.
In the example in James, et. al. at page 275, a natural cubic spline with 3 knots has 4 degrees of freedom, corresponding to K+1. However, I believe that this is wrong, and a natural cubic spline has K degrees of freedom, or in this exam question 4 degrees of freedom.

39. An actuary has a dataset with one dependent variable, Y, and five independent variables (X_1, X_2, X_3, X_4, X_5). She is trying to determine which subset of the predictors best fits the data, and is using a Forward Stepwise Selection procedure with no stopping rule. Below is a subset of the potential models:

Model	Dependent variable	RSS	Independent variable	p-value
1	Y	9,823	X_1	0.0430
			X_2	0.0096
2	Y	7,070	X_1	0.0464
			X_2	0.0183
			X_3	0.0456
3	Y	6,678	X_1	0.0412
			X_2	0.0138
			X_4	0.0254
4	Y	4,800	X_1	0.0444
			X_2	0.0548
			X_5	0.0254
5	Y	3,475	X_1	0.0333
			X_2	0.0214
			X_3	0.0098
			X_4	0.0274
			X_5	0.0076

The procedure just selected Model 1 as the new candidate model.
Determine which of the following independent variable(s) will be added to the model in the next iteration of this procedure.

A. No variables will be added

B. X_3 only

C. X_4 only

D. X_5 only

E. X_3, X_4 and X_5

39. D. Using the Forward Stepwise Selection procedure, we add one variable at each step.
So the next step cannot be Model 5.
We look for the smallest residual sum of squares (RSS) among Models 2 to 4.
Of these models, the best is Model 4, which adds $\mathbf{X_5}$.

Comment: When we are done, we will select a single best model from among the models selected at each stage, using cross-validated prediction error, C_p (AIC), or BIC, etc.

40. You have p = 10 independent variables and would like to select a linear model to fit the data using the following two procedures:

- Best Subset Selection (BSS)
- Forward Stepwise Selection (FSS)

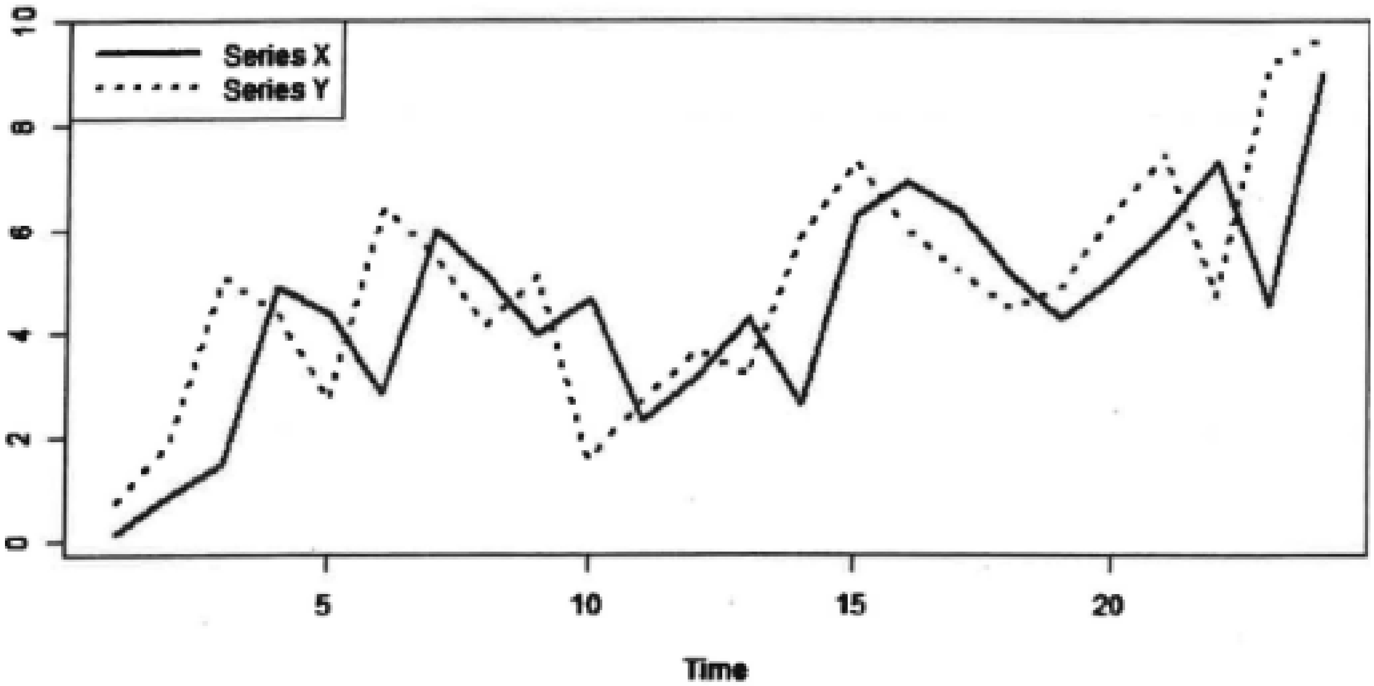
Let N_i be the maximum number of models fit by model selection procedure, i.

Calculate the ratio $\frac{N_{FSS}}{N_{BSS}}$.

A. Less than 0.005
B. At least 0.005, but less than 0.010
C. At least 0.010, but less than 0.050
D. At least 0.050, but less than 0.100
E. At least 0.100

40. D. For best subset selection, one fits all possible models; since each predictor can either be in or out of the model there are 2^p possible models. 2^{10} = 1024.
In general for forward stepwise selection, the number of models one has to fit is:
1 + p + p-1 + ... + 1 = 1 + p(p+1)/2. 1 + (10)(11)/2 = 56.
56/1024 = **0.0547**.
<u>Comment</u>: I am assuming that every model includes the intercept.
Forward Stepwise Selection has the advantage of fitting fewer models than Best Subset Selection; but at the cost of possibly missing the very best model.
Using Backward Stepwise Selection we would fit the same number of models as using Forward Stepwise Selection.

41. Two time series (X and Y) are shown in the graph below:



The autocorrelation variance function (acvf) and cross covariance function (ccvf) are estimated at lags 0, 1, 2 & 3 in the table below.

Lag	acvf(x)	acvf(y)	ccvf(x,y)
0	4.165	4.655	2.626
1	1.594	1.960	3.422
2	1.123	0.518	0.950
3	0.478	1.064	0.389

Calculate the sample lag 1 cross-correlation.
A. Less than 0.30
B. At least 0.30, but less than 0.55
C. At least 0.55, but less than 0.80
D. At least 0.80, but less than 1.05
E. At least 1.05

41. C. The estimated variances of x and y are the autocovariances for lag 0.

$\sigma_x = \sqrt{4.165} = 2.041$. $\sigma_y = \sqrt{4.655} = 2.158$.

$$r_1(x, y) = \frac{c_1(x,y)}{\sigma_x \sigma_y} = \frac{3.422}{(2.041)(2.158)} = \mathbf{0.777}.$$

Comment: $\gamma_k(x, y) = E[(x_{t+k} - \mu_x)(y_t - \mu_y)]$. $\gamma_1(x, y) = E[(x_{t+1} - \mu_x)(y_t - \mu_y)]$.

Based on the graph, for this time series, x_2 is similar to y_1, x_3 is similar to y_2, etc.

42. An actuary uses four separate models to fit a time series. All models have mean $\mu_X = 0$.

- Model 1: A random walk model with no drift
- Model 2: A stationary autoregressive process of order 1 with a root of the characteristic equation of the backwards shift operator equal to 3
- Model 3: A stationary autoregressive process of order 1 with a root of the characteristic equation of the backwards shift operator equal to 2
- Model 4: A non-stationary autoregressive process of order 1 with a root of the characteristic equation of the backwards shift operator greater than 0

The most recent values of x at time t, x_t, are given in the table below:

t	x_t
4	2
5	1
6	4

Determine which model will result in the smallest predicted values of x_7.

A. Model 1
B. Model 2
C. Model 3
D. Model 4
E. There is not enough information given to determine the correct answer.

42. B. Model 1: $\hat{x}_7 = x_6 = 4$.

Model 2: $\alpha = 1/3$. $\hat{x}_7 = (1/3)\, x_6 = 4/3$.

Model 3: $\alpha = 1/2$. $\hat{x}_7 = (1/2)\, x_6 = 4/2 = 2$.

Model 4: $\alpha \geq 1$. $\hat{x}_7 = \alpha\, x_6 \geq 4$.

The smallest predicted values of x_7 is for Model 2.

Comment: For an AR(1), the characteristic equation is: $1 - \alpha B = 0$, with root $1/\alpha$.
An autoregressive process is stationary if and only if all of the roots of its characteristic equation exceed one.

Thus a non-stationary autoregressive process of order one has $|\alpha| \geq 1$.

For a random walk, the predicted future value is equal to the last observed value.

43. You are given the following ARMA(p, q) model:

$$x_t = (\frac{3}{2})x_{t-1} - (\frac{1}{2})x_{t-2} + w_t - w_{t-1} + (\frac{1}{4})w_{t-2}$$

Determine (p, q) and whether the model is stationary and/or invertible.

A. p = 2, q = 2, Stationary, Invertible
B. p = 3, q = 3, Stationary, Not Invertible
C. p = 3, q = 3, Not Stationary, Not Invertible
D. p = 2, q = 2, Not Stationary, Invertible
E. p = 2, q = 2, Not Stationary, Not Invertible

43. D. Since the righthand side has a term involving x_{t-2} and a term involving w_{t-2}, this is an ARMA(2, 2) model.

$\theta_p(B) = 1 - \frac{3}{2}B + \frac{1}{2}B^2 = 0$. The roots are: $\frac{3/2 \pm \sqrt{(-3/2)^2 - (4)(1/2)(1)}}{(2)(1/2)} = 3/2 \pm 1/2 = 1$ or 2.

Since it is not the case that both roots exceed one in absolute value, this series is <u>not</u> stationary.

$\phi_q(B) = 1 - B + \frac{1}{4}B^2 = 0$. The roots are: $\frac{1 \pm \sqrt{(-1)^2 - (4)(1/4)(1)}}{(2)(1/4)} = 2 \pm 0 = 2$.

Since both roots exceed one in absolute value, this series is invertible.
<u>Comment</u>: One can find the roots by factoring the equations.

$1 - \frac{3}{2}B + \frac{1}{2}B^2 = (1 - B)(1- B/2)$. $1 - B + \frac{1}{4}B^2 = (1 - B/2)^2$.

44. You are given the following annual sales totals for a department store.

Year	Sales
2013	400
2014	375
2015	410
2016	420
2017	410
2018	525

Calculate the sample lag 2 autocorrelation.

A. Less than 0.00

B. At least 0.00. but less than 0.05

C. At least 0.05. but less than 0.10

D. At least 0.10. but less than 0.15

E. At least 0.15

44. B. The mean is 423.33.

$c_0 =$

$$\frac{(400-423.33)^2 + (375-423.33)^2 + (410-423.33)^2 + (420-423.33)^2 + (410-423.33)^2 + (525-423.33)^2}{6}$$

= 2264.

$c_2 =$

$$\frac{(400-423.33)(410-423.33) + (375-423.33)(420-423.33) + (410-423.33)(410-423.33) + (420-423.33)(525-423.33)}{6}$$

= 51.8.

The sample lag 2 autocorrelation is: c_2 / c_0 = 51.8/2264 = **0.023**.

45. A department store's annual sales, x_t, is modeled as an AR(3) process given by:

$$x_t = 450 - 0.9x_{t-1} + x_{t-3} + w_t$$

where w_t is white noise, with:

- $E(w_t) = 0$
- $Var(w_t) = \sigma^2$

You are given the following historical annual sales totals for this store:

Year	Sales
2013	400
2014	375
2015	410
2016	420
2017	410
2018	525

Calculate the two-step-ahead forecast value, x_{2020}.

A. Less than 400

B. At least 400, but less than 450

C. At least 450, but less than 500

D. At least 500, but less than 550

E. At least 550

45. D. $\hat{x}_{2019} = 450 - (0.9)(525) + 420 = 397.5$.

$\hat{x}_{2020} = 450 - (0.9)(397.5) + 410 =$ **502.25**.

Comment: The same data as in the previous question on this exam.

Similar to CAS S, 11/16, Q.44.

END OF EXAMINATION