Solutions to the Spring 2019 CAS Exam MAS-1

(Incorporating the Preliminary 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 3
2. Stochastic Models, Section 4
3. Stochastic Models, Section 10
4. Stochastic Models, Section 2
5. Loss and Frequency Distributions, Section 32
   (Can also be answered out of Stochastic Models, Section 2)

6. Stochastic Models, Section 2
7. Reliability Theory, Section 3
8. Stochastic Models, Section 18
9. Stochastic Models, Section 21
10. Stochastic Models, Section 26

11. Life Contingencies, Section 5
12. Life Contingencies, Section 5
13. Simulation, Section 4
14. Simulation, Section 12
15. Statistics, Section 5

16. Statistics, Section 33
17. Statistics, Section 33
18. Statistics, Section 5
19. Statistics, Section 38
20. Loss and Frequency Distributions, Section 38

21. Statistics, Section 23
22. Statistics, Section 23
23. Statistics, Section 26
24. Loss and Frequency Distributions, Section 18
25. Statistics, Section 31

26. Statistics, Section 31
27. Generalized Linear Models, Section 15
28. Statistical Learning, Section 5
29. Regression, Section 9
30. Regression, Section 4

[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. Regression, Section 9
33. Regression, Section 10
34. Statistical Learning, Section 2
35. Statistical Learning, Section 3

36. Regression, Section 18
37. Statistical Learning, Section 8
38. Statistical Learning, Section 15
39. Statistical Learning, Section 4
40. Statistical Learning, Section 13

41. Time Series, Section 2
42. Time Series, Section 3
43. Time Series, Section 4
44. Time Series, Section 16
45. Time Series, Section 8

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

	Number	Percent
Stochastic Models	8	17.8%
Loss & Freq. Dists.	3	6.7%
Statistics	10	22.2%
Regression	6	13.3%
GLMs	1	2.2%
Statistical Learning	7	15.6%
Life Contingencies	2	4.4%
Reliability	1	2.2%
Time Series	5	11.1%
Simulation	2	4.4%
Total	45	100.0%

1. Cars arrive according to a Poisson process at a rate of two per hour.
Calculate the probability that in a given hour, exactly one car will arrive during the first ten minutes and no other cars will arrive during that hour.

A. Less than 0.03
B. At least 0.03 but less than 0.06
C. At least 0.06 but less than 0.09
D. At least 0.09 but less than 0.12
E. At least 0.12

1. B. Number of cars over 10 minutes is Poisson with mean: (2)(10/60) = 1/3.

Prob[one car first ten minutes] = (1/3) $e^{-1/3}$ = 0.2388.

Number of cars over 50 minutes is Poisson with mean: (2)(50/60) = 5/3.

Prob[no car next fifty minutes] = $e^{-5/3}$ = 0.1889.

What happens over the first 10 minutes is independent of what happens over the last 50 minutes.
Prob[one car first ten minutes and no cars over the next 50 minutes] = (0.2388)(0.1889) = **0.0451**.

2. You are given the following information about the waiting time until a certain number of events occur:

- The underlying events follow a homogenous Poisson process
- T_n is the time until the n^{th} event occurs
- $E[T_2] = 2$

Calculate the variance of T_{10}.

A. Less than 4
B. At least 4, but less than 8
C. At least 8, but less than 12
D. At least 12, but less than 16
E. At least 16

2. C. $2 = E[T_2] = 2/\lambda$. $\Rightarrow \lambda = 1$.

T_{10} is the sum of 10 independent Exponential Distributions each with mean $1/\lambda = 1$.

Thus $\text{Var}[T_{10}] = (10)(1^2) =$ **10**.

Alternately, T_2 is Gamma with $\alpha = 2$ and θ. $\Rightarrow \alpha\,\theta = E[T_2] = 2$. $\Rightarrow \theta = 1$.

Thus T_{10} is Gamma with $\alpha = 10$ and $\theta = 1$. $\Rightarrow \text{Var}[T_{10}] = \alpha\,\theta^2 = (10)(1^2) =$ **10**.

3. You are given the following information:

- Accidents follow a compound Poisson process
- Accidents occur at the rate of $\lambda = 40$ per day
- Accident severity follows an exponential distribution with $\theta = 1{,}000$
- The insurance payment for each accident is subjected to a deductible
- V_1 is the variance of daily aggregate payments with a deductible of 100 per accident
- V_2 is the variance of daily aggregate payments with a deductible of 500 per accident

Calculate the ratio V_2 / V_1.

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

3. C. Due to the memoryless property of the Exponential, the non-zero payments follow the same Exponential Distribution with $\theta = 1{,}000$.

Thinning the Poisson, the number of non-zero payments is also Poisson with mean: S(d) 40.
Then the variance of the daily aggregate payments is:

S(d) 40 (second moment of the Exponential) = S(d) (40) (2) (1000^2) = 80 million S(d).

Thus $V_2 / V_1 = S(d_2) / S(d_1) = \exp[-500/1000] / \exp[-100/1000] = \exp[-0.4] =$ **0.670**.

4. You are given the following information:

- Computer lifetimes are exponentially distributed with mean of 10 months.
- Computer A has been functioning properly for 12 months.

Calculate the probability that Computer A will function properly for at least 4 more months.

A. Less than 0.64
B. At least 0.64, but less than 0.66
C. At least 0.66, but less than 0.68
D. At least 0.68, but less than 0.70
E. At least 0.70

4. C. Due to the memoryless property of the Exponential Distribution, it does not matter how long the computer has already been functioning.
S(4) = exp[-4/10] = **0.670**.

5. You are given the following information about random variable X:

- The hazard rate function is: $r(x) = \begin{cases} \dfrac{k^2}{3x}, \text{ for } x \geq 2 \\ 0, \text{ otherwise} \end{cases}$

- The value of the cumulative distribution function at x = 5 is:
  F(5) = 0.936.

Calculate the absolute value of k.

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

5. C. $\int_2^5 r(x)\, dx = \int_2^5 \frac{k^2}{3x}\, dx = (k^2/3)\{\ln(5) - \ln(2)\} = 0.3054\ k^2$.

Set this equal to: $-[\ln S(5)] = -\ln[1 - 0.936] = 2.749$.

$0.3054\ k^2 = 2.749. \Rightarrow k =$ **3.00**.

Comment: Similar to CAS3, 11/03, Q.19.

In general, $S(x) = \exp[-\int_0^x r(t)\, dt]. \Leftrightarrow -\ln[S(x)] = \int_0^x r(t)\, dt$.

For the Single Parameter Pareto Distribution, the hazard rate is: α/x. Also the support is $x \geq 2$.

Thus here we have a Single Parameter Pareto Distribution with $\theta = 2$ and $\alpha = k^2/3 = 3^2/3 = 3$.

$F(5) = 1 - (\theta/5)^\alpha = 1 - (3/5)^3 = 0..936$.

6. You are given the following information:

- The severity of each theft loss in a Homeowners insurance policy independently follows an exponential distribution with mean 2,000.
- The insurance company only pays the amount exceeding the per-loss deductible.
- σ_j is the standard deviation of the amount the insurance company pays per theft with a deductible of j.

Calculate the absolute difference between σ_{1000} and σ_{500}.

A. Less than 90
B. At least 90, but less than 100
C. At least 100, but less than 110
D. At least 110, but less than 120
E. At least 120

6. D. Due to the memoryless property of the Exponential, with a deductible, the amount the insurance company pays per non-zero claim is independent of the size of the deductible.
$\Rightarrow$ WIth a deductible of d, the payments per theft follow a two point mixture of zero and an Exponential with mean 2000, with weights F(d) and S(d).

This mixed distribution has mean S(d) 2000, second moment $(2)(2000^2)$ S(d),

and thus variance: (8 million) S(d) - (4 million) $S(d)^2$.

$S(500) = e^{-500/2000} = 0.77880.$

Thus $\sigma_{500} = \sqrt{(8 \text{ million})\ (0.77880) - (4 \text{ million})\ (0.77880^2)} = 1950.46.$

$S(1000) = e^{-1000/2000} = 0.60653.$

Thus $\sigma_{1000} = \sqrt{(8 \text{ million})\ (0.60653) - (4 \text{ million})\ (0.60653^2)} = 1838.67.$

The absolute difference between σ_{1000} and σ_{500} is: $|1950.46 - 1838.67|$ = **111.8**.

Alternately, wIth a deductible of d, the payments per theft follow an aggregate distribution with a Bernoulli frequency with mean S(d), and an Exponential Severity with mean 2000.

This has variance: S(d) (2000^2) + (2000^2) S(d){1 - S(d)} = (8 million) S(d) - (4 million) $S(d)^2$.

Proceed as before.

Comment: Difficult.
See Example 5.4 in Introduction to Probability Models by Ross.
The moments of a mixture are the mixture of the moments.
Variance of an aggregate distribution is:

(mean frequency) (variance of severity) + (mean severity)^2 (variance of frequency).

7. You are given the following information:

- In a toolbox there are two types of components that all perform the same function:
  - There are 4 components of type A, each with reliability of 0.600
  - There are 20 components of type B, each with reliability of 0.300
  - All components are independent
- Using only the components in this toolbox, you want to construct a parallel system with a reliability of at least 0.995

Calculate the minimum number of components needed to create this system.

A. Fewer than 3

B. At least 3, but fewer than 5

C. At least 5, but fewer than 7

D. At least 7, but fewer than 9

E. At least 9

7. E. One wishes to use first the more reliable components of type A.

Using in parallel all 4 components of type A, the chance of failure is: $(1 - 0.6)^4 = 2.56\%$.

This is too big, so we need to add in parallel b components of type B.

The chance of failure is now: $(0.0256)(1 - 0.3)^b$.

Set this equal to the desired chance of failure of 0.005.

$0.005 = (0.0256)(1 - 0.3)^b$. $\Rightarrow$ $b = \ln(0.1953) / \ln(0.7) = 4.58$.

Thus we need at least 5 components of type B, for a total of at least **9** components.

Comment: With 9 components, the chance of failure is: $(0.4^4)(0.7^5) = 0.00430$.

With 8 components, the chance of failure is: $(0.4^4)(0.7^4) = 0.00615$.

Adding components to a parallel system increases the reliability, while adding components to a series system decreases the reliability.

8. On any given day, Alan will play either one or two games of billiards.

- If he plays one game today, he'll play one game tomorrow with probability 0.6
- If he plays two games today, he'll play two games tomorrow with probability 0.7
- His probability of winning each game is 0.6

It is now Monday and Alan played two games today.
Calculate Alan's expected number of wins two days from now on Wednesday.
A. Less than 0.90
B. At least 0.90, but less than 0.95
C. At least 0.95, but less than 1.00
D. At least 1.00, but less than 1.05
E. At least 1.05

8. C. The transition matrix is: $\begin{array}{c} \text{one} \\ \text{two} \end{array} \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix}$.

(0, 1) P = (0.3, 0.7). (0.3, 0.7) P = (0.39, 0.61).
Thus the average number of games he plays on Wednesday is: (0.39)(1) + (0.61)(2) = 1.61.
The average number of games he wins on Wednesday is: (0.6)(1.61) = **0.966**.

9. You are given the following information about a 40-state Markov chain:

- The states are numbered: {0, 1, ... , 39}
- $P_{m,n}$ is the one-step transition probability from State m to State n
- $P_{0,0} = \frac{1}{2}$
- $P_{0,i} = \frac{1}{78}$ for i = 1, 2, ..., 39
- $P_{i,0} = \frac{1}{39}$ for i = 1, 2, ..., 39
- $P_{i,\, i+1} = \frac{38}{39}$ for i = 1, 2, ..., 38
- $P_{39,1} = \frac{38}{39}$

Calculate the long run probability of being in State 0.

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

9. A. $P = \begin{pmatrix} 1/2 & 1/78 & 1/78 & \cdots & 1/78 & 1/78 \\ 1/39 & 0 & 38/39 & \cdots & 0 & 0 \\ 1/39 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1/39 & 0 & 0 & \cdots & 0 & 38/39 \\ 1/39 & 38/39 & 0 & \cdots & 0 & 0 \end{pmatrix}$.

Let π the vector of long run probabilities.
We want π P = π , and that the probabilities sum to one.
The first equation is: $\pi_0/2 + \pi_1/39 + \pi_2/39 + ... + \pi_{39}/39 = \pi_0$.
Since the probabilities add to one: $\pi_1 + \pi_2 + ... + \pi_{39} = 1 - \pi_0$.
$\Rightarrow \pi_0/2 + (1 - \pi_0)/39 = \pi_0$. $\Rightarrow 39\pi_0 + 2 - 2\pi_0 = 78\pi_0$. $\Rightarrow \pi_0 =$ **2/41** = **0.0488**.
Comment: Difficult under exam conditions.
The second equation is: $\pi_0/78 + \pi_{39}\,38/39 = \pi_1$.
The third equation is: $\pi_0/78 + \pi_1\,38/39 = \pi_2$.
The last equation is: $\pi_0/78 + \pi_{38}\,38/39 = \pi_{39}$.
By symmetry: $\pi_1 = \pi_2 = ... = \pi_{39}$.
Thus, since the probabilities add to one: $\pi_1 = \pi_2 = ... = \pi_{39} = 1/41$.

10. You are given the following information about two gamblers:

- The two gamblers start with a total fortune of 100 tokens between them
- They plan to go to two different and independent casinos to try their luck
- Each gambler has a 50% chance of winning 1 token and a 50% chance of losing 1 token at each play
- All plays are mutually independent
- The gamblers initially divide their 100 combined tokens to maximize the probability that both of their fortunes will eventually reach 100 tokens

Calculate the product of their initial fortunes, in tokens.

A. Less than 1200
B. At least 1200, but less than 1600
C. At least 1600, but less than 2000
D. At least 2000, but less than 2400
E. At least 2400

10. E. Since the game is fair, given that a gambler starts with x tokens, the chance that he will reach 100 tokens before going broke is: x/100.
One gambler starts with x and the other gambler takes the remaining 100 - x tokens.
Thus the product of their probabilities of each succeeding is:

$$\frac{x}{100}\ \frac{100 - x}{100} = \frac{100x - x^2}{10{,}000}.$$

This function is zero at x = 0 and zero at x = 100, so we can maximize it by setting the derivative equal to zero.

$$0 = \frac{100 - 2x}{10{,}000} \Rightarrow 100 - 2x = 0. \Rightarrow x = 50.$$

$\Rightarrow$ Each gambler takes 50 tokens. (50)(50) = **2500**.
<u>Comment</u>: Difficult. Splitting the tokens equally would have been a reasonable guess.
When the tokens are split equally, the chance of both gamblers succeeding is: (1/2)(1/2) = 1/4.

11. You are given the following information:

- An annuity-due is issued to a 30-year-old that will pay 1 each year until either she dies or reaches age 50, whichever comes fIrst.
- Mortality follows the Illustrative Life Table.
- Annual interest rate i = 6%.

Calculate the actuarial present value of this annuity.

A. Less than 10.0
B. At least 10.0, but less than 10.5
C. At least 10.5, but less than 11.0
D. At least 11.0, but less than 11.5
E. At least 11.5

11. E. There are at most 20 payments.

$\ddot{a}_{30:\overline{20|}} = \ddot{a}_{30} - {}_{20}p_{30}\ \ddot{a}_{50}\ v^{20}$ = 15.8561 - (8,950,901/ 9,501,381)(13.2668)/1.06^{20} = **11.9591**.

Comment: $\ddot{a}_{x:\overline{n|}} = \ddot{a}_x - {}_np_x\ \ddot{a}_{x+n}\ v^n$.

We need to eliminate the payments that would otherwise have been made at age 50 and beyond by subtracting $\ddot{a}_{50}$; however, we first need to discount $\ddot{a}_{50}$ for both mortality and interest.

12. You are given the following information for a whole life annuity-due of 10,000 on (x) that is payable annually:

- $q_x = 0.010$
- $q_{x+1} = 0.012$
- Interest rate $i = 0.06$
- $\ddot{a}_{x+1} = 10.905$

Calculate the absolute change in the actuarial present value of this annuity due if p_{x+1} is decreased by 0.01.

A. Less than 900
B. At least 900, but less than 925
C. At least 925, but less than 950
D. At least 950, but less than 975
E. At least 975

12. C. $10.905 = \ddot{a}_{x+1} = 1 + p_{x+1}\, v + p_{x+1}\, p_{x+2}\, v^2 + ...$

$\Rightarrow\ 9.905 = p_{x+1}\, v + p_{x+1}\, p_{x+2}\, v^2 + ... = p_{x+1}\, (v + p_{x+2}\, v^2 + ...)$

Now if p_{x+1} is changed from 0.988 to 0.978, then:

$p_{x+1}\, (v + p_{x+2}\, v^2 + ...) = (0.978/0.988)\, (9.905) = 9.8047.\ \Rightarrow \ddot{a}_{x+1} = 1 + 9.8047 = 10.8047.$

Thus the absolute change in $\ddot{a}_{x+1}$ is: $|10.8047 - 10.905| = 0.1003$.

$\ddot{a}_x = 1 + v\, p_x\, \ddot{a}_{x+1} = 1 + (0.990/1.06)\ddot{a}_{x+1}$.

Thus the absolute change in $10{,}000\ddot{a}_x$ is:

(10,000) (0.990/1.06) (0.1003) = **937**.

Alternately, originally: $10{,}000\ \ddot{a}_x = (10{,}000)\ (1 + v\, p_x\, \ddot{a}_{x+1})$

$= (10{,}000)\ \{1 + (0.990/1.06)\ (10.905)\} = 111{,}848.6.$

After the change: $10{,}000\ \ddot{a}_x = (10{,}000)\ (1 + v\, p_x\, \ddot{a}_{x+1})$

$= (10{,}000)\ \{1 + (0.990/1.06)\ (10.8047)\} = 110{,}911.8.$

Thus the absolute change in $10{,}000\ddot{a}_x$ is:

$|110{,}911.8 - 111{,}848.6| =$ **937**.

<u>Comment</u>: When p_{x+1} is decreased, the average lifetime decreases as does the value of an annuity.

13. You use the inversion method to simulate two random numbers X from a Weibull (θ , τ) distribution using two independent draws U from a uniform distribution on (0,1).
The results of this simulation are shown below:

i	X_i	U_i
1	416.28	0.5
2	357.36	0.4

Calculate the value of θ used to perform this simulation.

A. Less than 425
B. At least 425, but less than 475
C. At least 475, but less than 525
D. At least 525, but less than 575
E. At least 575

13. C. $1 - \exp[-(416.28/\theta)^\tau] = 0.5. \Rightarrow (416.28/\theta)^\tau = 0.6931.$

$1 - \exp[-(357.36/\theta)^\tau] = 0.4. \Rightarrow (357.36/\theta)^\tau = 0.5108.$

Dividing the two equations: $1.1649^\tau = 1.3569. \Rightarrow \tau = \ln(1.3569) / \ln(1.1649) = 2.$

$\Rightarrow (357.36/\theta)^2 = 0.5108. \Rightarrow \theta =$ **500**.

Comment: Equivalent to fitting a Weibull Distribution via Percentile Matching, as discussed in "Mahler's Guide to Statistics".

Check: $1 - \exp[-(416.28/500)^2] = 0.500.$ $1 - \exp[-(357.36/500)^2] = 0.400.$

14. You are given the following simulation process to generate random variable X using the rejection method:

- X has a beta distribution, f(x), with parameters a = 2, b = 10, θ = 1, and is concentrated on the interval (0, 1).
- The rejection method uses g(x) = 1, for 0 < x < 1.
- The rejection procedure is as follows:
  - Step 1: Generate independent random numbers Y and U, which are both uniform on (0,1).
  - Step 2: If $U \le \frac{f(Y)}{c\, g(Y)}$ stop and set X = Y. OtherwIse return to Step 1.

Calculate the minimum possible value for c.

A. Less than 3.0

B. At least 3.0, but less than 3.5

C. At least 3.5, but less than 4.0

D. At least 4.0, but less than 4.5

E. At least 4.5

14. D. $c = \frac{(a+b-1)!}{(a-1)!\,(b-1)!} \left(\frac{a-1}{a+b-2}\right)^{a-1} \left(\frac{b-1}{a+b-2}\right)^{b-1} = \frac{(2+10-1)!}{(2-1)!\,(10-1)!} \left(\frac{2-1}{2+10-2}\right)^{2-1} \left(\frac{10-1}{2+10-2}\right)^{10-1} =$

$\frac{11!}{1!\,9!} \left(\frac{1}{10}\right)^{1} \left(\frac{9}{10}\right)^{9}$ = **4.262**.

Alternately, $f(x) = \frac{\Gamma(12)}{\Gamma(2)\,\Gamma(10)}\, x^{2-1}\, (1-x)^{10-1} = \frac{11!}{1!\,9!}\, x\, (1-x)^9 = 110\, x\, (1-x)^9$.

We wish to maximize: $f(x) / g(x) = 110\, x\, (1-x)^9$.

Set the derivative with respect to x equal to zero: $110(1-x)^9 - 990x(1-x)^8 = 0$.

$\Rightarrow (1-x) = 9x. \Rightarrow x = 1/10$.

$f(0.1) / g(0.1) = (110)\,(0.1)\,(0.9^9) =$ **4.262**.

<u>Comment</u>: For a Beta Distribution with θ = 1, $f(x) \le c\, g(x)$, with g being the uniform distribution on (0, 1) and $c = \frac{(a+b-1)!}{(a-1)!\,(b-1)!} \left(\frac{a-1}{a+b-2}\right)^{a-1} \left(\frac{b-1}{a+b-2}\right)^{b-1}$.

15. You are given the following information:

- $x_1, x_2, \ldots, x_{500}$ is a random sample of 500 claims
- The distribution function is $F(x) = 1 - \left(\frac{200}{x}\right)^{\alpha}$; for $x > 200$ and $\alpha > 0$
- $\sum_{i=1}^{500} x_i = 117{,}378{,}845$ and $\sum_{i=1}^{500} \ln(x_i) = 3{,}646$
- The parameter α is estimated using the method of maximum likelihood

Calculate the estimated probability that a claim will exceed 400.

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

15. D. $f(x) = \alpha\, 200^{\alpha} / x^{\alpha+1}$. $\ln f x) = \ln(\alpha) + \alpha \ln(200) - (\alpha+1)\ln(x)$.

loglikelihood is: $500 \ln(\alpha) + 500\, \alpha \ln(200) - (\alpha+1) \sum_{i=1}^{500} \ln(x_i)$.

Set the partial derivative of the loglikelihood equal to zero:

$$0 = 500/\alpha + 500 \ln(200) - \sum_{i=1}^{500} \ln(x_i).$$

$\Rightarrow \alpha = 500 / \{3646 - 500 \ln(200)\} = 1/\{7.292 - \ln(200)\} = 0.5016.$

$S(400) = (200/400)^{0.5016} =$ **0.706**.

<u>Comment</u>: For the Single Parameter Pareto Distribution fit via maximum likelihood:

$$\alpha = \frac{N}{\sum \ln[x_i / \theta]} = \frac{N}{\sum \ln[x_i] - N\theta}.$$

16. Suppose that $X_1, \ldots, X_{10}$ is a random sample from the standard normal distribution, N(0, 1).

An actuary is using $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ to estimate the population variance, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Calculate the bias of this estimator.

A. Less than -0.2
B. At least -0.2, but less than 0.0
C. At least 0.0, but less than 0.2
D. At least 0.2, but less than 0.4
E. At least 0.4

16. B. Since the sample variance is an unbiased estimator of the variance:

$$1 = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]. \Rightarrow E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = n - 1 = 10 - 1 = 9.$$

$$\Rightarrow E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = E\left[\frac{1}{10}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{1}{10}\ 9 = 0.9.$$

Thus, the bias = 0.9 - 1 = **-0.1**.

Comment: The estimator is biased downwards; the bias is negative.

17. You are given the following information:

- τ is a parameter of a distribution
- The true value of τ is 2
- $\hat{\alpha}$ and $\hat{\beta}$ are two uncorrelated estimators for τ
- $E(\hat{\alpha}) = 2.0$
- $Var(\hat{\alpha}) = 5.0$
- $E(\hat{\beta}) = 3.0$
- $Var(\hat{\beta}) = 1.0$

Consider the class of estimators of τ which are of the form: $w\hat{\alpha} + (1 - w)\hat{\beta}$.

Calculate the value of w that results in an estimator with the smallest mean squared error.

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

17. B. $E[w\hat{\alpha} + (1 - w)\hat{\beta}] = w\,E[\hat{\alpha}] + (1 - w)E[\hat{\beta}] = w2 + (1-w)3 = 3 - w$.

The true value is 2. $\Rightarrow$ Bias = (3 - w) - 2 = 1 - w.

$Var[w\hat{\alpha} + (1 - w)\hat{\beta}] = w^2\,Var[\hat{\alpha}] + (1 - w)^2\,Var[\hat{\beta}] = w^2\,(5) + (1 - w)^2\,(1) = 6w^2 - 2w + 1$.

MSE = Variance + Bias^2 = $6w^2 - 2w + 1 + (1-w)^2 = 7w^2 - 4w + 2$.

Minimize the mean squared error by setting the derivative equal to zero:
$0 = 14w - 4$. $\Rightarrow$ w = 4/14 = **0.286**.

18. You are given a random sample from a uniform distribution [0, θ]:

1.3, 2.7, 8.2, 5.3, X

You find that the estimate of θ using method of moments is equal to the estimate of θ using the maximum likelihood.

Calculate the largest possible value of X.

A. Less than 4

B. At least 4, but less than 6

C. At least 6, but less than 8

D. At least 8, but less than 10

E. At least 10

18. E. For maximum likelihood: $\hat{\theta}$ = maximum of the sample = 8.2 or X.

For the method of moments: θ /2 = (1.3 + 2.7 + 8.2 + 5.3 + X)/5. $\Rightarrow \hat{\theta}$ = 7 + 0.4X.

For X ≤ 8.2, we want: 8.2 = 7 + 0.4X. $\Rightarrow$ X = 3.

For X > 8.2, we want: X = 7 + 0.4X. $\Rightarrow$ X = **11.67**.

Comment: For some samples, the Method of Moments applied to the uniform distribution could result in an estimate of θ that is less than the maximum of the sample; this is impossible.

For example, for the sample {1, 1, 1, 1, 10}: $\hat{\theta}$ = (2)(14/5) = 5.6 < 10.

19. We observe a random sample, $\{X_1, X_2, ..., X_n\}$ from a continuous uniform distribution $U(0, \beta)$.

$\hat{\beta} = \max(X_1, X_2, ..., X_n)$ is chosen as an estimator of β.

I. $\hat{\beta}$ is both the maximum likelihood estimator and method of moments estimator of β

II. $\hat{\beta}$ is sufficient statistic for β

III. The minimum variance unbiased estimator of β is a function of $\hat{\beta}$

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)

19. D. For maximum likelihood, $\hat{\beta}$ = maximum of the sample.

For method of moments, (estimated β)/2 = average.

So the two estimators are not equal in general. Statement I is not true.
Since the maximum of the sample is the maximum likelihood estimator, it is a sufficient statistic.
Alternately, let I[y] be zero if $y < 0$ and 1 if $y \geq 0$.

Then the likelihood is: $I[\beta - X_{(n)}] / \beta^n$. Thus $X_{(n)} = \max(X_1, X_2, ..., X_n)$ is a sufficient statistic for β.

Statement II is true.

$E[\hat{\beta}] = \beta\, n/(n+1)$. Thus $\hat{\beta}\ (n+1)/n$ is an unbiased estimator of β.

It is in fact the minimum variance unbiased estimator of β. Thus Statement III is true

<u>Comment</u>: Exercise 7.4.2 in Hogg, McKean, and Craig shows that for the uniform distribution $U(0, \beta)$, the minimum variance unbiased estimator of β is a function of the maximum.

This uses Theorem 7.4.1 (Lehmann and Scheffe) in Hogg, McKean, and Craig: If there is a complete sufficient statistic of β, and there is an unbiased estimator of β, then this estimator is the unique MVUE of β.

For an exponential family, if one can find an unbiased estimator that is function of a sufficient statistic, then this is a Minimum Variance Unbiased Estimator (MVUE), however; the uniform distribution is not an exponential family, so this result does not apply here.

If a minimum variance unbiased estimator (MVUE) of β exists, and β has a sufficient statistic, then the MVUE is a function of the sufficient statistic; however, we would need to establish that an MVUE exists in this case of a uniform distribution.

The maximum follows a Beta Distribution with a = n, b = 1, and $\theta = \beta$, with variance:

$$\theta^2 \frac{a\, b}{(a+b)^2\, (a+b+1)} = \beta^2 \frac{n}{(n+1)^2\, (n+2)}.$$

Therefore, since the MVUE is $\hat{\beta}\ (n+1)/n$, the corresponding minimum variance is:

$$\beta^2 \frac{n}{(n+1)^2\, (n+2)} \left(\frac{n+1}{n}\right)^2 = \beta^2 \frac{1}{n\,(n+2)}.$$

For some samples, the maximum likelihood estimator and method of moments estimators are equal, as shown in the previous question on this exam.

20. A sample of losses has the following observations:

{400, 600, 900, 1100, 1200}

A rectangular kernel with bandwidth 300 is used to estimate the probability density function (pdf) of the losses at x = 800. The resulting estimate is $\tilde{f}_R(800)$.

A triangular kernel with bandwidth 300 is used to estimate the pdf of the losses at x = 800.

The resulting estimate is $\tilde{f}_T(800)$.

The absolute difference in the estimates is denoted z, where $z = \left|\tilde{f}_R(800) - \tilde{f}_T(800)\right|$.

Calculate z.

A. Less than 0.0001

B. At least 0.0001, but less than 0.0002

C. At least 0.0002, but less than 0.0003

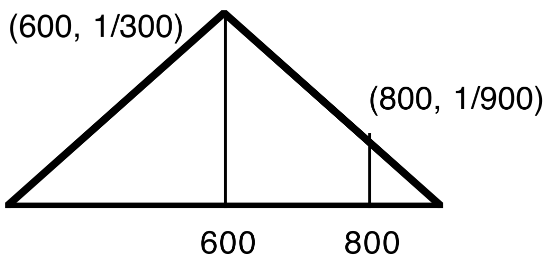
D. At least 0.0003, but less than 0.0004

E. At least 0.0004

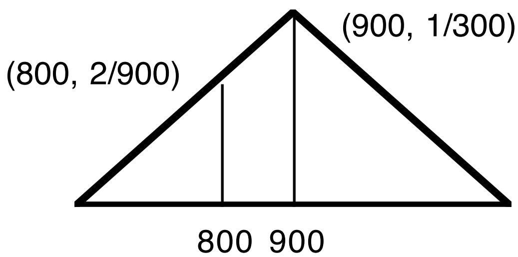
20. D. For both the rectangular and triangular kernel, only points within 300 of 800 will contribute.
There are three points within 300 of 800. (The endpoints are included, so 1100 contributes.)
The rectangular kernel has height 1/(2b) = 1/600, and width 2b.

Thus, since 3 out of 5 data points are close enough to 800: $\tilde{f}_R(800) = (3/5)(1/600) = 0.001000$.

The triangular kernel has height 1/b = 1/300, and width 2b.



The triangular kernel centered at 600 contributes: (1/3)(1/300).



The triangular kernel centered at 900 contributes: (2/3)(1/300).
The triangular kernel centered at 1100 contributes: (0)(1/300).

Thus, $\tilde{f}_T(800) = (1/5)(1/300)(1/3 + 2/3 + 0) = 0.000667$.

$\left|\tilde{f}_R(800) - \tilde{f}_T(800)\right| = |0.001000 - 0.000667| = \mathbf{0.00033}$.

<u>Comment</u>: All kernels have an area of one.

21. $X_1, X_2, ..., X_{100}$ is a random sample from a Bernoulli distribution with probability of success q.
You perform the hypothesis test:

H_0: q = 0.5 vs H_1: q > 0.5

The critical region is chosen to be C = $\left(\sum_{i=1}^{100} X_i \geq 60 \right)$.

Calculate the probability of a Type I error.

A. Less than 0.01
B. At least 0.01, but less than 0.02
C. At least 0.02, but less than 0.03
D. At least 0.03, but less than 0.04
E. At least 0.04

21. C. The sum of 100 independent Bernoullis is a Binomial with m = 100.

Prob[Type I error] = Prob[reject given H_0 is true] = Prob[$\sum_{i=1}^{100} X_i \geq 60$ | q = 0.5].

A Binomial Distribution with m = 100 and q = 0.5 has mean of 50 and standard deviation of:
$\sqrt{(100)(0.5)(1 - 0.5)} = 5$.

Thus using the Normal Approximation with continuity correction:

Prob[$\sum_{i=1}^{100} X_i \geq 60$ | q = 0.5] $\cong 1 - \Phi\left[\frac{59.5 - 50}{5}\right] = 1 - \Phi[1.9] = 1 - 0.9713 =$ **2.87%**.

<u>Comment</u>: Using a computer, the exact answer is 2.8444%.

22. Let X be a Bernoulli random variable with probability of success q.
You want to perform the following hypothesis test:

- H_0: q = 0.75
- H_1: q > 0.75

We will reject H_0 in favor of H_1 if $\sum_{i=1}^{30} X_i \geq k$.

Calculate the smallest k such that significance level of this test will be at most 0.02.
A. Fewer than 27
B. 27
C. 28
D. 29
E. At least 30

22. C. The sum is Binomial with m = 30.
For q = 0.75, the mean is: (0.75)(30) = 22.5, and the variance is: (0.75)(0.25)(30) = 5.625.

$$0.02 = \text{significance level} = \text{Prob}[\sum_{i=1}^{30} X_i \geq k \mid q = 0.75] = 1 - \Phi\left[\frac{k - 0.5 - 22.5}{\sqrt{5.625}}\right].$$

$$\Rightarrow \frac{k - 0.5 - 22.5}{\sqrt{5.625}} = 2.06. \Rightarrow k = 27.89. \Rightarrow \text{Smallest k is } \mathbf{28}.$$

Alternately, we can determine the densities of a Binomial Distribution with m = 30 and q = 0.75.
f(30) = 0.00018. f(29) = 0.0018. f(28) = 0.00863. f(27) = 0.02685.

Thus $\text{Prob}[\sum_{i=1}^{30} X_i \geq 28] = 1.06\% < 2\%$. $\text{Prob}[\sum_{i=1}^{30} X_i \geq 27] = 3.74\% > 2\%$.

$\Rightarrow$ Smallest k is **28**.

23. Two independent random samples of sizes 10 and 12 are taken from two separate normal distributions with unknown variances σ_1^2 and σ_2^2, respectively.

- S_1^2 and S_2^2 are the unbiased sample variances
- $S_1^2 = 4$
- You want to perform the following hypothesis test:
  - $H_0: 2\sigma_1^2 = \sigma_2^2$
  - $H_1: 2\sigma_1^2 < \sigma_2^2$
- The p-value of this test is 1.0%

Calculate S_2^2.

A. Less than 10
B. At least 10, but less than 20
C. At least 20, but less than 30
D. At least 30, but less than 40
E. At least 40

23. E. $\frac{S_2^2}{2S_1^2}$ follows an F-Distribution with 12-1 = 11 and 10-1 = 9 degrees of freedom.

From the Table, the 1% critical value is 5.178.

$\Rightarrow 5.178 = \frac{S_2^2}{2S_1^2} = \frac{S_2^2}{(2)(4)}$. $\Rightarrow S_2^2 = \mathbf{41.424}$.

Comment: Put the larger value in the alternative hypothesis in the numerator of the F-statistic.

24. You are given the following information:

- Random variable X has an exponential distribution
- $\text{Var}(X) = \frac{1}{9}$

Calculate the median of the distribution.

A. Less than 0.15
B. At least 0.15, but less than 0.20
C. At least 0.20, but less than 0.25
D. At least 0.25, but less than 0.30
E. At least 0.30

24. C. For an Exponential with mean θ: $E[X^2] = 2\theta^2$, and $\text{Var}[X] = \theta^2$.

$\theta^2 = 1/9. \Rightarrow \theta = 1/3.$

Set $0.5 = 1 - \exp[-3x]. \Rightarrow x = \ln(2)/3 =$ **0.231**.

<u>Comment</u>: The median of an Exponential Distribution is: $\theta \ln(2)$.

25. A type of machine has a lifetime following an exponential distribution with mean μ.
All machine lifetimes are independent.
To estimate the expected lifetime of this machine, the factory measures lifetimes of 4 sample machines and takes the average excluding the lowest and highest values. Denote this estimate $\hat{\mu}$.

Calculate the ratio $E(\hat{\mu})/\mu$.

A. Less than 0.8
B. At least 0.8, but less than 1.0
C. At least 1.0, but less than 1.2
D. At least 1.2, but less than 1.4
E. At least 1.4

25. B. For a sample of size N from an Exponential Distribution with mean θ:

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

Thus the expected values of the four order statistics of the Exponential are:
$\mu/4$, $\mu(1/4 + 1/3)$, $\mu(1/4 + 1/3 + 1/2)$, and $\mu(1/4 + 1/3 + 1/2 + 1)$.
$E[X_{(2)}] = \mu\ 7/12$. $E[X_{(3)}] = \mu\ 13/12$.

Thus $E(\hat{\mu}) = (\mu\ 7/12 + \mu\ 13/12)/2 = \mu\ 10/12$. $\Rightarrow$ $E(\hat{\mu})/\mu = 5/6 =$ **0.833**.

Comment: Similar to CAS S, 11/16, Q.24.
Excluding the high and low produces a biased estimator of the mean;
however, it can be shown that this is an asymptotically unbiased estimator.
$E[X_{(4)}] = \mu\ 25/12$. $E[X_{(1)} + X_{(2)} + X_{(3)} + X_{(4)}] = \mu\ 48/12 = 4\mu$.
Thus the average of all of the order statistics would be an unbiased estimator of the mean;
this is true in general.

26. For a general liability policy, loss amounts, Y, follow the Weibull distribution with probability density function:

$$f(y) = \exp\left[-\left(\frac{y}{\theta}\right)^2\right] \left(\frac{2}{y}\right) \left(\frac{y}{\theta}\right)^2, \ y > 0$$

For reinsurance purposes we are interested in the distribution of the largest loss amount in a random sample of size 10, denoted by $Y_{(10)}$.

The 90th percentile of $Y_{(10)}$ is equal to $k\theta$ for some constant k.

Calculate k.

A. Less than 1.2
B. At least 1.2, but less than 1.6
C. At least 1.6, but less than 2.0
D. At least' 2.0, but less than 2.4
E. At least 2.4

26. D. This is a Weibull Distribution with $\tau = 2$. $F(y) = 1 - \exp\left[-\left(\frac{y}{\theta}\right)^2\right]$.

Therefore, the distribution function of the maximum of a sample of size 10 is: $\left\{1 - \exp\left[-\left(\frac{y}{\theta}\right)^2\right]\right\}^{10}$.

90th percentile of $Y_{(10)}$ is where this distribution function is 0.9.

$$0.9 = \left\{1 - \exp\left[-\left(\frac{y}{\theta}\right)^2\right]\right\}^{10}. \Rightarrow 1 - \exp\left[-\left(\frac{y}{\theta}\right)^2\right] = 0.98952. \Rightarrow \exp\left[-\left(\frac{y}{\theta}\right)^2\right] = 0.01048.$$

$$\Rightarrow \left(\frac{y}{\theta}\right)^2 = 4.558. \Rightarrow y/\theta = 2.135. \Rightarrow y = 2.135\theta. \Rightarrow k = \mathbf{2.135}.$$

Comment: Similar to CAS ST, 11/15, Q.18.

27. A statistician uses a logistic model to predict the probability of success, π , of a binomial random variable.

You are given the following information:

- There is one predictor variable, X, and an intercept in the model
- The estimates of π at x = 4 and 6 are 0.88877 and 0.96562, respectively

Calculate the estimated intercept coefficient, b_0, in the logistic model that produced the above probability estimates.

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

27. B. $0.88877 = \frac{\exp[\beta_0 + 4\beta_1]}{1 + \exp[\beta_0 + 4\beta_1]}$. $\Rightarrow$ $\beta_0 + 4\beta_1 = \ln(\frac{0.88877}{1 - 0.88877}) = 2.0782$.

$0.96562 = \frac{\exp[\beta_0 + 6\beta_1]}{1 + \exp[\beta_0 + 6\beta_1]}$. $\Rightarrow$ $\beta_0 + 6\beta_1 = \ln(\frac{0.96562}{1 - 0.96562}) = 3.3353$.

Subtracting the two equations: $2\beta_1 = 1.2571$. $\Rightarrow$ $\beta_1 = 0.6285$. $\Rightarrow$ $\beta_0 =$ **-0.4359**.

28. Determine which one of the following statements about ridge regression is false.

A. As the tuning parameter $\lambda \to \infty$, the coefficients tend to zero.

B. The ridge regression coefficients can be calculated by determining the coefficients

$$\hat{\beta}_1^R, \hat{\beta}_2^R, ..., \hat{\beta}_3^R \text{ that minimize: } \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{n} \beta_j x_{ij})^2 + \lambda \sum_{i=1}^{p} \beta_i{}^2$$

C. Unlike standard least squares coefficients, ridge regression coefficients are not scale equivariant.

D. The shrinkage penalty is applied to all the coefficient estimates except for the intercept.

E. Ridge regression shrinks the coefficient estimates, which has the benefit of reducing the bias.

28. E. Ridge regression shrinks the coefficient estimates, as does as the Lasso, which has the benefit of reducing the <u>variance</u>, but at the cost of <u>increasing</u> the bias.

<u>Comment</u>: There is a trade-off between the variance and the bias; we prefer the smallest Test MSE. See Figure 6.5 in an <u>Introduction to Statistical Learning</u>.

As lambda increases, there is less flexibility and thus more bias.

Statements A, C, and D are also true of the Lasso.

In the case of the Lasso, as lambda increases some of the coefficients become zero.

29. Tim uses an ordinary least squares regression model to predict salary based on Experience and Gender. Gender is a qualitative variable and is coded as follows:

$$\text{Gender} = \begin{cases} 1 & \text{if Male} \\ 0 & \text{if Female} \end{cases}$$

His analysis results in the following output:

Coefficients	Estimate	Std. Error	t-value	Pr(>ltl)
Intercept	18169.300	212.2080	85.62027	2.05E-14
Experience	1110.233	59.8224	18.55881	1.75E-08
Gender	169.550	162.9177	10.38285	2.62E-06

Abby uses the same data set but codes the Gender as follows:

$$\text{Gender} = \begin{cases} 1 & \text{if Female} \\ 0 & \text{if Male} \end{cases}$$

Calculate the value of the Intercept in Abby's model.

A. At most 18,169.3
B. Greater than 18,169.3, but at most 18,400.0
C. Greater than 18,400.0, but at most 18,600.0
D. Greater than 18,600.0
E. The answer cannot be computed from the information given

29. B. Tim and Abby should produce the same estimates.

For a female with no experience, Tim's estimate is: 18,169.3 + 169.5 = 18,338.85.

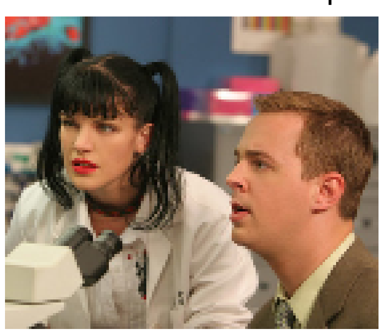
For a female with no experience, Abby's estimate is: β_0. ⇒ β_0 = **18,338.85**.

Comment: For a male with no experience, Tim's estimate is: 18,169.3.

For a female with no experience, Abby's estimate is: 18,338.85 + β_2.

⇒ 18,169.3 = 18,338.85 + β_2. ⇒ β_2 = -169.55. This is Abby's coefficient for Gender.

Tim and Abby should have the same fitted coefficient for Experience.



30. You are given the following information about a linear model:

• $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

Observed	Estimated
Y's	Y's
2.441	1.827
3.627	3.816
5.126	5.806
7.266	7.796
10.570	9.785

• Residual Sum of Squares = 1.772

Calculate the R^2 of this model.

A. Less than 0.6
B. At least 0.6, but less than 0.7
C. At least 0.7, but less than 0.8
D. At least 0.8, but less than 0.9
E. At least 0.9

30. E. The mean of the observed Y's is 5.806.
The total sum of squares is:
$(2.441 - 5.806)^2 + (3.627 - 5.806)^2 + (5.126 - 5.806)^2 + (7.266 - 5.806)^2 + (10.570 - 5.806)^2 =$
41.361.

Thus $R^2 = 1 - \frac{\text{Residual Sum of Squares}}{\text{Total Sum of Squares}} = 1 - 1.772/41.361 =$ **0.957**.

Alternately, the model sum of squares is:
$(1.827 - 5.806)^2 + (3.816 - 5.806)^2 + (5.806 - 5.806)^2 + (7.796 - 5.806)^2 + (9.785 - 5.806)^2 =$
39.585.
$R^2 = 39.585/41.361 =$ **0.957**.
<u>Comment</u>: The residual sum of squares can be computed:
$(2.441 - 1.827)^2 + (3.627 - 3.816)^2 + (5.126 - 5.806)^2 + (7.266 - 7.796)^2 + (10.570 - 9.785)^2$
$= 1.772$.
The averages of the observed and estimated Y's are equal; this is true in general for multiple regressions that include an intercept. Also $\hat{Y} - \bar{Y}$ and $Y - \hat{Y}$ are uncorrelated; this is true in general for multiple regressions that include an intercept.

31. 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:

$$X = \begin{pmatrix} 1 & 0 & 1 & 9 \\ 1 & 1 & 1 & 15 \\ 1 & 1 & 1 & 8 \\ 0 & 1 & 1 & 7 \\ 0 & 1 & 1 & 6 \\ 0 & 0 & 1 & 6 \end{pmatrix}; \qquad y = \begin{pmatrix} 21 \\ 32 \\ 19 \\ 17 \\ 15 \\ 15 \end{pmatrix}; \qquad X^T X = \begin{pmatrix} 3 & 2 & 3 & 32 \\ 2 & 4 & 4 & 36 \\ 3 & 4 & 6 & 51 \\ 32 & 36 & 51 & 491 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 1.38 & 0.25 & 0.54 & -0.16 \\ 0.25 & 0.84 & -0.20 & -0.06 \\ 0.54 & -0.20 & 1.75 & -0.20 \\ -0.16 & -0.06 & -0.20 & 0.04 \end{pmatrix}$$

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

$$(X^T X)^{-1} X^T y = \begin{pmatrix} 0.297 \\ -0.032 \\ 3.943 \\ 1.854 \end{pmatrix}; \qquad X(X^T X)^{-1} X^T y = \begin{pmatrix} 20.93 \\ 32.03 \\ 19.04 \\ 16.89 \\ 15.04 \\ 15.07 \end{pmatrix}; \qquad \sigma^2 = 0.012657$$

Calculate how many observations are influential, using a unity threshold for Cook's distance.

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

31. B. The values of $\hat{y}$ are given by: $X(X^TX)^{-1} X^Ty$ = (20.93, 32.03, 19.04, 16.89, 15.04, 15.07).

Thus the residuals are: $y - \hat{y}$ =

(21, 32, 19, 17, 15, 15) - (20.93, 32.03, 19.04, 16.89, 15.04, 15.07) =
(0.07, -0.03, -0.04, 0.11, -0.04, -0.07).
The diagonal elements of the Hat matrix are the leverage values, h_{ii}:

(0.684, 0.975, 0.797, 0.418, 0.443, 0.684).

We get the standardized residuals by dividing by σ = 0.1125 and by $\sqrt{1 - h_{ii}}$:

$$\left(\frac{0.07}{0.1125\sqrt{1-0.684}}, \frac{-0.03}{0.1125\sqrt{1-0.975}}, \frac{-0.04}{0.1125\sqrt{1-0.797}}, \frac{0.11}{0.1125\sqrt{1-0.418}},\right.$$

$$\left.\frac{-0.04}{0.1125\sqrt{1-0.443}}, \frac{-0.07}{0.1125\sqrt{1-0.684}}\right) = (1.11, -1.69, -0.79, 1.28, -0.48, -1.11).$$

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

$$(1.11, -1.69, -0.79, 1.28, -0.48, -1.11)^2 \frac{1}{3+1} \frac{(0.684, 0.975, 0.797, 0.418, 0.443, 0.684)}{\{1 - (0.684, 0.975, 0.797, 0.418, 0.443, 0.684)\}} =$$

(0.67, <u>27.8</u>, 0.61, 0.29, 0.05, 0.67).
Only **one** data point has a Cook's Distance greater than 1.
<u>Comment</u>: Long and difficult.
Since the design matrix has six rows, there are six observations.
Since the design matrix has four columns, there are three slopes in addition to the intercept.
The computed values of Cook's distance were affected by the limited number of significant digits shown in the question. Using a computer to fit the multiple regression, the values of Cook's Distance are: 0.6534, 19.25, 0.7537, 0.2817, 0.0407, 0.6534.
Out of 6 data points, it would be unusual for more than one to be influential; thus A or B are good guesses.
The Hat Matrix, H, is as usual symmetric, with number of rows nd columns equal to the number of observations, six. One could verify that H is idempotent: $H^2 = H$.
As is the case when there is an intercept, each row and each column of H adds to one.
The trace of H, in other words the sum of the leverages along the diagonal of H, is four, the number of parameters (3 slopes plus an intercept.)

32. 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:

$$X = \begin{pmatrix} 1 & 0 & 1 & 9 \\ 1 & 1 & 1 & 15 \\ 1 & 1 & 1 & 8 \\ 0 & 1 & 1 & 7 \\ 0 & 1 & 1 & 6 \\ 0 & 0 & 1 & 6 \end{pmatrix}; \qquad y = \begin{pmatrix} 21 \\ 32 \\ 19 \\ 17 \\ 15 \\ 15 \end{pmatrix}; \qquad X^T X = \begin{pmatrix} 3 & 2 & 3 & 32 \\ 2 & 4 & 4 & 36 \\ 3 & 4 & 6 & 51 \\ 32 & 36 & 51 & 491 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 1.38 & 0.25 & 0.54 & -0.16 \\ 0.25 & 0.84 & -0.20 & -0.06 \\ 0.54 & -0.20 & 1.75 & -0.20 \\ -0.16 & -0.06 & -0.20 & 0.04 \end{pmatrix}$$

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

$$(X^T X)^{-1} X^T y = \begin{pmatrix} 0.297 \\ -0.032 \\ 3.943 \\ 1.854 \end{pmatrix}; \qquad X(X^T X)^{-1} X^T y = \begin{pmatrix} 20.93 \\ 32.03 \\ 19.04 \\ 16.89 \\ 15.04 \\ 15.07 \end{pmatrix}; \qquad \sigma^2 = 0.012657$$

Calculate the modeled estimate of the intercept parameter.

A. Less than 0
B. At least 0, but less than 1
C. At least 1, but less than 2
D. At least 2, but less than 3
E. At least 3

32. E. The third column of the design matrix X is all ones, so this corresponds to the intercept.
$(X^T X)^{-1} X^T y$ is the column of fitted coefficients; the third element is the fitted intercept: **3.943**.
Comment: Usually the first column of the design matrix is all ones, corresponding to the intercept.
Uses the same setup as the previous question on this exam.

33. You are given the following information regarding a GLM that was used to predict claim severity of Homeowners insurance losses:

- The available predictors are:
  - Number of Stories of home (" 1" or "2+")
  - Square Footage of the home ("Under 1500", "1501 to 2500", or "Over 2500")
- The partial ANOVA results, which are available to evaluate whether the interaction term "Number of Stories x Square Footage" is predictive, are given below:

Response variable	Insured Loss
Response distribution	Normal
Link	Identity

Parameters	Degrees of freedom	Sum of squares
Number of Stories, Square Footage, and intercept	4	1,128.19
Improvement from Interaction: Number of Stories x Square Footage	2	1.25
Residuals	6	

- The scaled deviance of the model without the interaction terms is 2.64

Calculate the F-statistic that is used to evaluate whether the interaction term is significant.

A. Less than 2.00

B. At least 2.00, but less than 2.25

C. At least 2.25, but less than 2.50

D. At least 2.50, but less than 2.75

E. At least 2.75

33. D. A GLM with an identity link and a Normal response is a multiple regression.
The residual sum of squares for the model without interactions is the scaled deviance of 2.64.
Thus the residual sum of squares for the model with interactions is: 2.64 - 1.25 = 1.39.
As shown in the row for residuals, this has 6 degrees of freedom. ⇒ The denominator is: 1.39/6.
The model with interactions improves the sum of squares by 1.25 and adds two degrees of freedom. ⇒ The numerator is: 1.25/2.
⇒ F = (1.25/2) / (1.39/6) = **2.698**, with 2 and 6 degrees of freedom.
Comment: Difficult. See Tables 6.1 and 6.5, as well as page 83, in Dobson and Barnett.
While it is less than clear, this is the usual test of nested models:

$$\text{F-Statistic} = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(N-k-1)} = (1.25/2) / (1.39/6) = 2.698.$$

The restricted model is without interactions, while the unrestricted model includes interactions.
Residual Sum of Squares for the Restricted Model = 2.64.
Since $RSS_R - RSS_{UR}$ is given as 1.25:
Residual Sum of Squares for the Unrestricted Model = 2.64 - 1.25 = 1.39.
q = the dimension of the restriction = the difference of the number of parameters of the models = 2.
We are given N - k - 1 = 6.
k is the number of slopes in the unrestricted model, which in this case is:
one for number of stories, two for square footage, and two for interactions, totaling 5.
Including the intercept, the unrestricted model has 6 degrees of freedom;
this is two more degrees of freedom than the 4 shown for the restricted model.
Thus N - 6 = 6. ⇒ N = 12. There are 12 data points.

Parameters	Degrees of freedom	Sum of squares	Mean Square
Number of Stories, Square Footage, and intercept	4	1,128.19	
Improvement from Interaction: Number of Stories x Square Footage	2	1.25	1.25/2 = 0.625
Residual	6	1.39	1.39/6 = 0.23167
F-Statistic			0.625 / 0.23167 = 2.698.

For 2 and 6 degrees of freedom, the 20% critical value is 2.130, while the 10% critical value is 3.463.
Since 2.130 < 2.698 < 3.463, the p-value is between 20% and 10%.
Using a computer, the p-value is 14.6%.
1,128.19 shown in the first row is the Model Sum of Squares for the model without interactions.

34. A statistician has a dataset with n = 50 observations and p = 22 independent predictors.
He is using 10-fold cross validation to select from a variety of available models.
Calculate the number of times that the first observation will be included in the training dataset as part of this procedure.

A. 0
B. At least 1, but less than 5
C. At least 5, but less than 10
D. At least 10, but less than 20
E. At least 20

34. C. Using 10-fold cross validation, the data is split into 10 subsets, and each subset in turn acts as the test data, while the remaining 9 subsets act as the training data.
Each data point is in the test set once and in the training set the remaining 10 - 1 = **9** times.
Comment: In the case of Leave-One-Out Cross-Validation, in turn each data point forms a test set by itself. Thus in this situation, each data point would be in the training data 50 - 1 = 49 times.

35. You are given the following information about a linear model:

- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
- Three bootstrap samples were drawn from the data

Bootstrap Sample	Estimate for β_0
1	10.525
2	2.499
3	16.456

Calculate the standard error of these bootstrap estimates of β_0.

A. Less than 6
B. At least 6, but less than 8
C. At least 8, but less than 10
D. At least 10, but less than 12
E. At least 12

35. B. The mean of the three estimates is 9.8267.
The sample variance of the three estimates is:
$\{(10.525 - 9.8267)^2 + (2.499 - 9.8267)^2 + (16.456 - 9.8267)^2\} / (3 - 1) = 49.0652$.

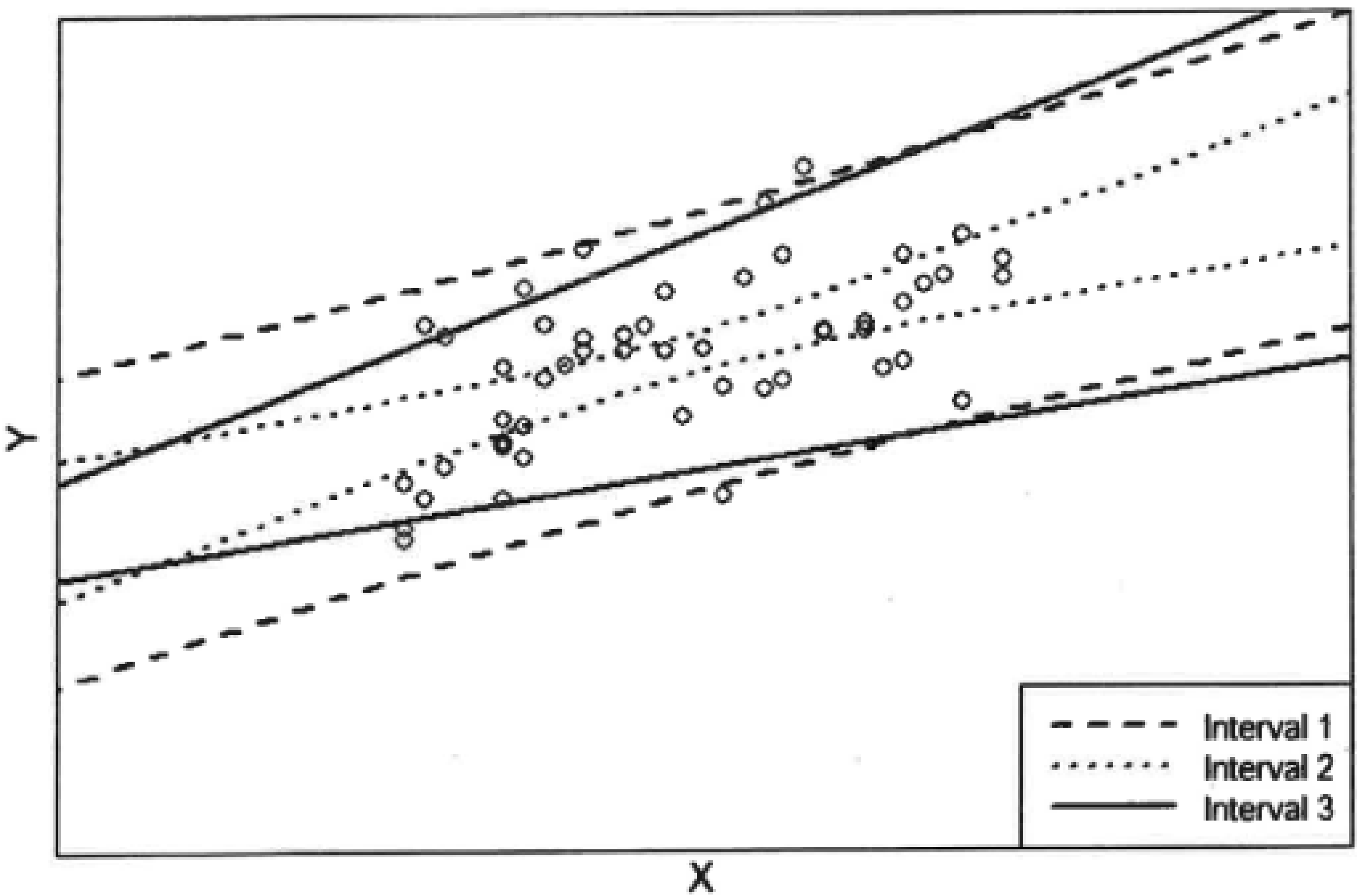
Thus, the standard error of these bootstrap estimates of β_0 is: $\sqrt{49.0652} = \mathbf{7.005}$.

<u>Comment</u>: In practical applications, one would simulate many more bootstrap samples.
For multiple linear regression, one can estimate the standard error of the intercept without using bootstrapping; however, this would depend on the assumptions underlying multiple regression, while the bootstrap estimates do not.

36. An ordinary least squares regression model is fit with the following model form:

$$E[Y_i] = \beta_0 + \beta_1 X_i.$$

After fitting the model, the following plot with the original data (points) and three sets of 95% intervals are provided:



Let "CI" be the 95% confidence interval for $E(Y_i)$, and let "PI" be the 95% prediction interval for Y_i.

Determine which one of the following best describes the intervals shown above.

A. Interval 1 = CI Interval 2 = PI
B. Interval 1 = PI Interval 2 = CI
C. Interval 1 = CI Interval 3 = PI
D. Interval 1 = PI Interval 3 = CI
E. None of (A), (B), (C), or (D) are correct

36. B. Both types of intervals are centered on the fitted straight line, and get wider as one gets further in either direction from the mean of X. This eliminates Interval 3.
The 95% prediction interval for Y_i is wider than the 95% confidence interval $E(Y_i)$.

Thus **Interval 1 = PI** and **Interval 2 = CI**.
<u>Comment</u>: I do not know what type of interval if any that the third interval is meant to be.

37. You want to perform a regression of Y onto predictors X_1, X_2, ... , X_p, using a large number of observations, and are considering the following modelling techniques:

- Lasso Regression
- Partial Least Squares
- Principal Component Analysis
- Ridge Regression

Determine how many of the above modelling procedures perform variable selection.

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

37. B. Only the Lasso performs variable selection.

<u>Comment</u>: The textbook refers to "the lasso" rather than "lasso regression".

For the Lasso, as the tuning parameter lambda approaches infinity, some coefficients become and stay zero. While for Ridge Regression as the tuning parameter lambda approaches infinity, some coefficients approach zero, they do not become and stay zero.

If one used only the first principal component, then one has decided to use some linear combination of the original variables.

38. Determine which one of the following statements regarding Generalized Additive Models (GAMs) is false.

A. Natural Splines, Regression Splines, Smoothing Splines, Local Regression, Polynomial Regression, and Step Functions are all types of models that can be used as building blocks for GAMs.

B. A limitation of GAMs is that interactions cannot be added to the model.

C. The smoothness of a continuous predictor variable can be summarized via degrees of freedom.

D. GAMs are a useful representation if we are interested in inference, since you can examine the effect of the predictor variables on the response while holding all of the other predictor variables constant.

E. GAMs allow for non-linear relationships between each predictor variable and the response.

38. B or C. (see Comment)

The smoothness of each function on a predictor, $f_j(X_j)$, can be summarized via degrees of freedom.

Thus Statement C is not true.

The main limitation of GAMs is that the model is restricted to be additive.

However, we can add interaction terms by including additional predictors such as $X_j X_k$.

In addition we could add low-dimensional interaction functions of the form $f_{jk}(X_j, X_k)$ into the model.

Thus Statement B is not true.

Comment: The CAS preliminary answer key has B.

Statement E is the basic reason for using Generalized Additive Models.

$f_j(X_j)$ are the building blocks of GAMs. $y = \beta_0 + f_1(x_1) + f_2(x_2) + ... + f_p(x_p) + \varepsilon$.

See pages 285 to 286 of An Introduction to Statistical Learning.

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

Below are the results for all possible models:

Model	# of Non Intercept Parameters	Parameters	R^2	AIC
1	0	I	0	1.9
2	1	I, 1	0.56	1.4
3	1	I, 2	0.57	1.2
4	1	I, 3	0.55	1.6
5	1	I, 4	0.52	1.7
6	1	I, 5	0.51	1.8
7	2	I, 1, 2	0.61	1.0
8	2	I, 1, 3	0.64	0.5
9	2	I, 1, 4	0.63	0.8
10	2	I, 1, 5	0.69	0.0
11	2	I, 2, 3	0.61	1.0
12	2	I, 2, 4	0.62	0.9
13	2	I, 2, 5	0.68	0.2
14	2	I, 3, 4	0.66	0.4
15	2	I, 3, 5	0.64	0.5
16	2	I, 4, 5	0.60	1.1

Model	# of Non Intercept Parameters	Parameters	R^2	AIC
17	3	I, 1, 2, 3	0.73	1.3
18	3	I, 1, 2, 4	0.71	1.5
19	3	I, 1, 2, 5	0.72	1.4
20	3	I, 1, 3, 4	0.75	1.0
21	3	I, 1, 3, 5	0.76	0.8
22	3	I, 1, 4, 5	0.79	0.2
23	3	I, 2, 3, 4	0.78	0.6
24	3	I, 2, 3, 5	0.74	1.2
25	3	I, 2, 4, 5	0.75	1.1
26	3	I, 3, 4, 5	0.73	1.3
27	4	I, 1, 2, 3, 4	0.88	1.6
28	4	I, 1, 2, 3, 5	0.80	2.1
29	4	I, 1, 2, 4, 5	0.87	1.8
30	4	I, 1, 3, 4, 5	0.83	2.0
31	4	I, 2, 3, 4, 5	0.85	1.9
32	5	I, 1, 2, 3, 4, 5	0.90	3.5

• AICj is the AIC of the model chosen by Actuary j

Determine the correct ordering of the AIC values of the three selected models.

A. $AIC_A < AIC_B < AIC_C$

B. $AIC_A = AIC_B < AIC_C$

C. $AIC_A < AIC_C < AIC_B$

D. $AIC_A = AIC_C < AIC_B$

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

39. A. For Best Subset Selection, we need only consider for each number of parameters the model with the best (largest) R^2.

Model	Number of Parameters	AIC
1	1	1.9
3	2	1.2
10	3	**0**
22	4	0.2
27	5	1.6
32	6	3.5

Model 10 has the best (smallest) AIC at 0. $AIC_A = 0$.

For Forward Stepwise Regression, we start with the model with just an intercept, and then at each stage we choose the best model (biggest R^2) that adds one predictor to the current model.

Model	Parameters	Number of Parameters	AIC
1	I	1	1.9
3	I, 2	2	1.2
13	I, 2, 5	3	**0.2**
25	I, 2, 4, 5	4	1.1
29	I, 1, 2, 4, 5	5	1.8
32	I, 1, 2, 3, 4, 5	6	3.5

Model 13 has the best (smallest) AIC at 0.2. $AIC_B = 0.2$.

For Backwards Stepwise Regression, we start with the model using all predictors, and then at each stage we choose the best model (biggest R^2) that subtracts one predictor from the current model.

Model	Parameters	Number of Parameters	AIC
32	I, 1, 2, 3, 4, 5	6	3.5
27	I, 1, 2, 3, 4	5	1.6
23	I, 2, 3, 4	4	0.6
14	I, 3, 4	3	**0.4**
4	I, 3	2	1.6
1	I	1	1.9

Model 14 has the best (smallest) AIC at 0.4. $AIC_C = 0.4$.

$\mathbf{AIC_A < AIC_B < AIC_C}$.

<u>Comment</u>: Very similar to MAS-1, 11/18, Q.39.

40. You are fitting a model to predict the height of an adult male using shoe size as a predictor.
You have a sample size of 200.
Determine which one of the following model forms will both have a discontinuity in the fitted curve and most likely overfit the data.
A. Piecewise cubic with two knots
B. Piecewise linear with two knots
C. Cubic spline with one knot
D. Natural cubic spline with two knots
E. Smoothing spline with 10 degrees of freedom

40. A. Piecewise linear and piecewise cubic models are each discontinuous.
The piecewise cubic has more parameters than the piecewise linear model, and is thus more likely to overfit the data.
Comment: Cubic spline, natural cubic spline, and smoothing spline, are each continuous.
The piecewise cubic model with two knots has 4 parameters each on 3 intervals,
for total of 12 parameters.
The piecewise linear model with two knots has 2 parameters each on 3 intervals,
for total of 6 parameters.

41. The time series below shows study manuals sold by month by company XYZ:

t	Month	Year	Study Manuals Sold		t	Month	Year	Study Manuals Sold
1	Jan	2016	52		13	Jan	2017	60
2	Feb	2016	18		14	Feb	2017	22
3	Mar	2016	16		15	Mar	2017	17
4	Apr	2016	14		16	Apr	2017	19
5	May	2016	16		17	May	2017	21
6	Jun	2016	16		18	Jun	2017	23
7	Jul	2016	18		19	Jul	2017	25
8	Aug	2016	50		20	Aug	2017	60
9	Sep	2016	20		21	Sep	2017	28
10	Oct	2016	16		22	Oct	2017	22
11	Nov	2016	18		23	Nov	2017	24
12	Dec	2016	20		24	Dec	2017	28

Calculate the monthly additive effect at time t = 13 using the centering approach.

A. Less than 25

B. At least 25, but less than 30

C. At least 30, but less than 35

D. At least 35, but less than 40

E. At least 40

41. C. We need to take the annual average centered at time 13, $\hat{m}_{13}$:

$\{18/2 + 50 + 20 + 16 + 18 + 20 + 60 + 22 + 17 + 19 + 21 + 23 + 25/2\} / 12 = 25.625$.

$\hat{s}_{13} = x_{13} - \hat{m}_{13} = 60 - 25.625 = \mathbf{34.375}$.

Comment: See Equation 1.6 in Introductory Times Series with R.

With a longer time series, we would get more than one estimate of the additive monthly effect of January, and would average these estimates.

The monthly multiplicative effect at time t = 13 is: 60/25.65 = 2.341.

42. Consider the following time-series data for price of a stock on January 1 for the last 5 years:

Date	Jan 1, 2013	Jan 1, 2014	Jan 1, 2015	Jan 1, 2016	Jan 1, 2017
Price	63.18	81.89	103.43	123.90	133.53

Calculate the sample autocorrelation at lag 1 for this data.

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

42. C. The sample mean is 101.186.

$$c_k = \frac{\sum_{i=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{n}.$$

The sample autocovariance function for lag 0 is c_0:

$\{(63.18 - 101.186)^2 + (81.89 - 101.186)^2 + (103.43 - 101.186)^2 + (123.90 - 101.186)^2$
$+ (133.53 - 101.186)^2\} / 5 = 676.777.$

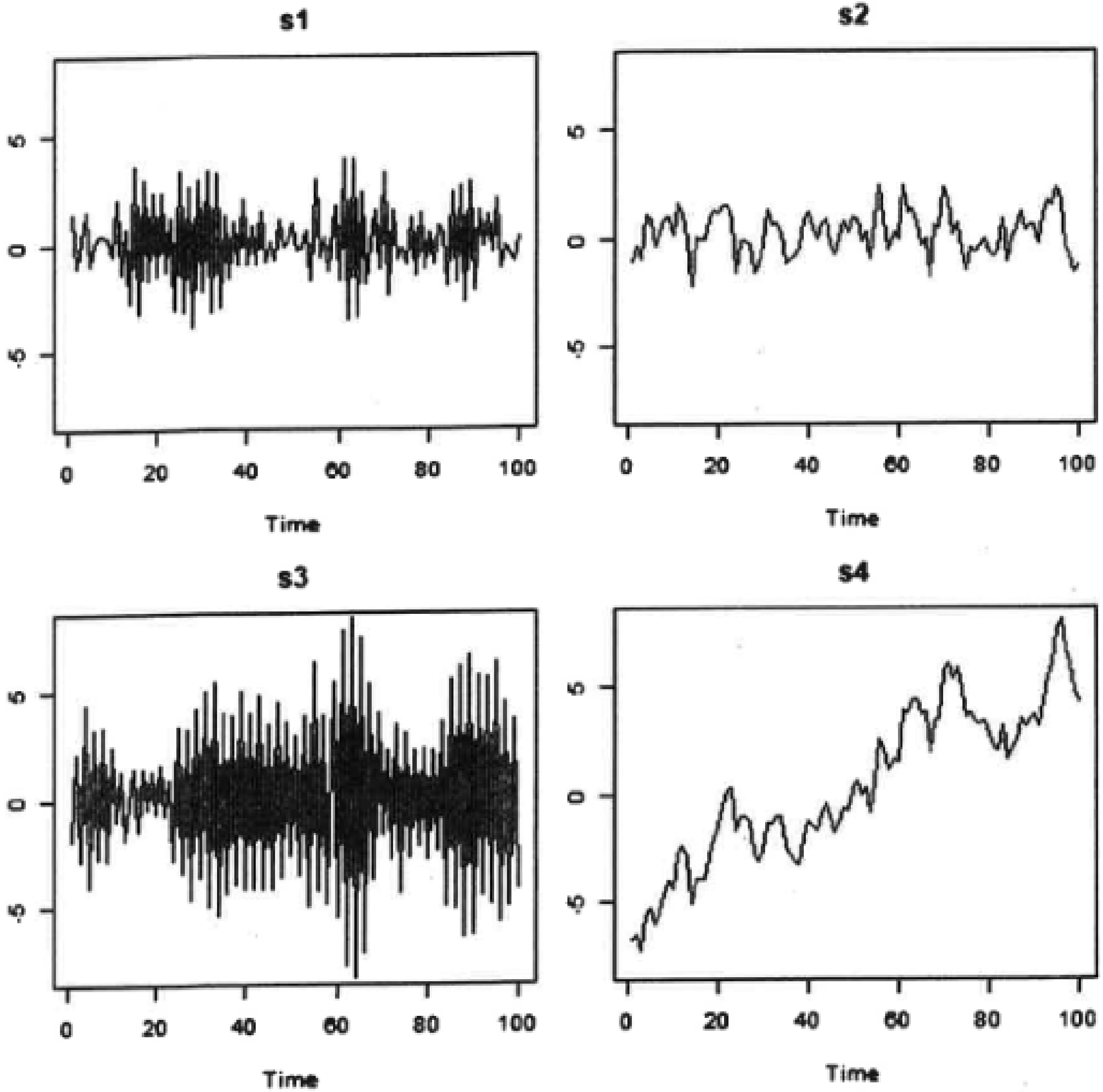
The sample autocovariance function for lag 1 is c_1:

$\{(63.18 - 101.186)(81.89 - 101.186) + (81.89 - 101.186)(103.43 - 101.186)$
$+ (103.43 - 101.186)(123.90 - 101.186) + (123.90 - 101.186)(133.53 - 101.186)\} / 5$
$= 295.139.$

The sample autocorrelation for lag 1 is: c_1/c_0 = 295.139/676.777 = **0.436**.

43. You are given an autoregressive time series of order 1: $x_t = \alpha_1 x_{t-1} + w_t$

and the following time series graphs:



Determine the most likely coefficient α_1 for each graph, s1-s4.

A. s1 = 0.90, s2 = 0.50, s3 = 0.99, s4 = 0.99
B. s1 = 0.99, s2 = -0.50, s3 = 0.90, s4 = -0.99
C. s1 = -0.90, s2 = 0.50, s3 = -0.99, s4 = 0.99
D. s1 = -0.99, s2 = 0.50, s3 = -0.90, s4 = 0.99
E. s1 = -0.90, s2 = 0.99, s3 = -0.99, s4 = 0.50

43. C. For an AR(1) model: $\rho_k = \alpha^k$. In particular, $\rho_1 = \alpha$.

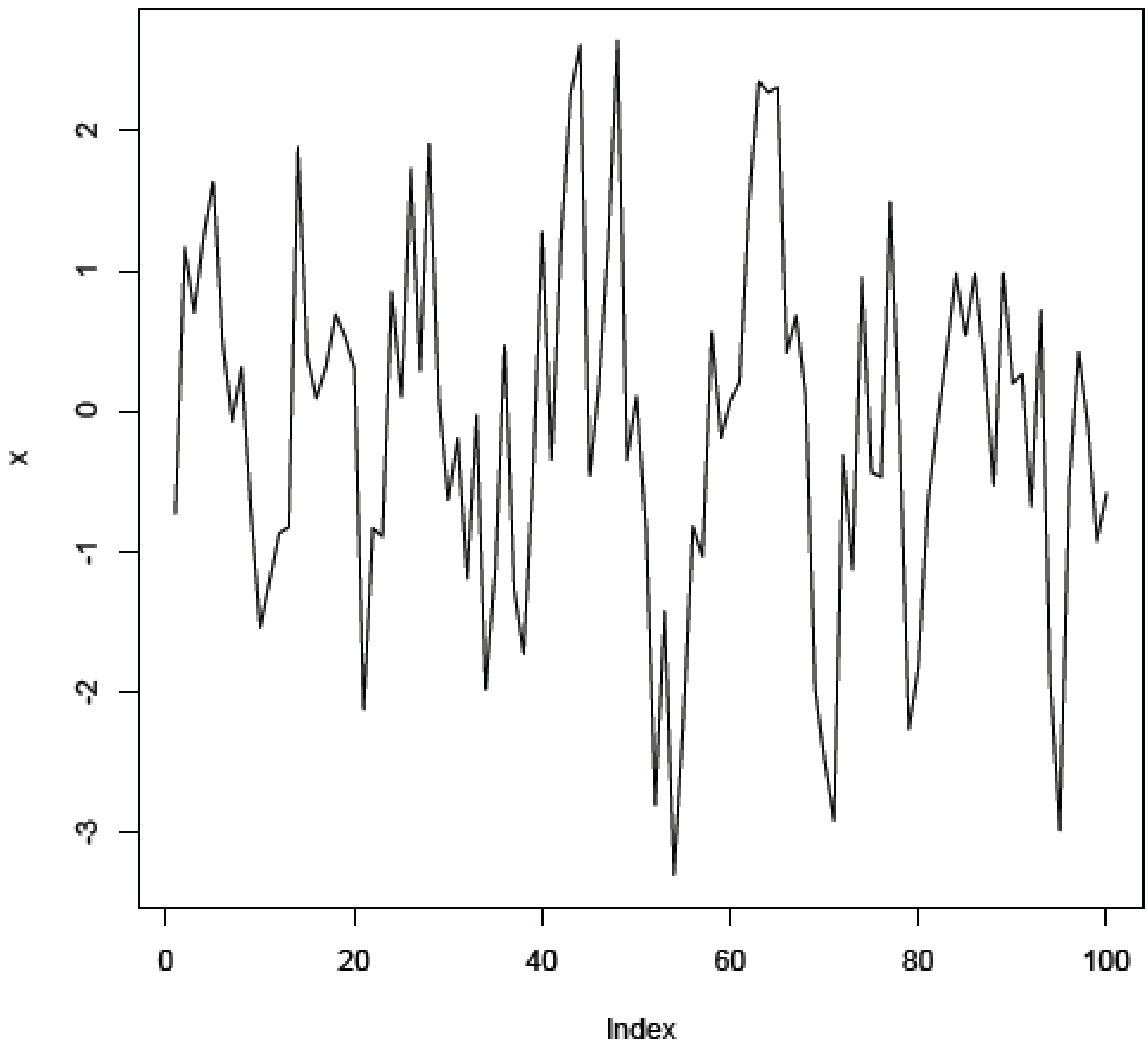
Graphs 1 and 3 have negative lag 1 autocorrelations; x_t and x_{t-1} are similar in absolute value but of different sign. Thus $\alpha_1 < 0$. Graph 3 jumps back and forth more than than in graph 1.

Consistent with this, graph 1 could have α_1 = -0.99, while graph 1 has α_1 = -0.90.

Graphs 2 and 4 have positive lag 1 autocorrelations; x_t and x_{t-1} are similar in absolute value and of the same sign. Thus $\alpha_1 > 0$. In graph 2, the correlation of successive values is much less than in graph 4. Consistent with this, graph 2 could have α_1 = 0.50, while graph 4 has α_1 = 0.99.

<u>Comment</u>: Correlograms would be more useful than graphs of the values of the series over time.

A simulation of an AR(1) model with α_1 = 0.50 and σ^2 = 1:



44. You are given the following statements regarding deterministic and stochastic trends:

I. Deterministic trends can be modeled using regression.

II. Stochastic trends usually have a plausible, physical explanations such as an increase in population.

III. Short term extrapolation of deterministic trends can be justified by claiming that underlying trends will usually change slowly in comparison with the forecast lead time.

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)

44. C. Statements I and III are true. Statements II would be true of instead deterministic trends.

Comment: See page 91 of Introductory Time Series with R:

"The practical difference between stochastic and deterministic trends is that we extrapolate the latter when we make forecasts. We justify short-term extrapolation of deterministic trends by claiming that underlying trends will usually change slowly in comparison with the forecast lead time. For the same reason, short-term extrapolation should be based on a line, maybe fitted to the most recent data only, rather than a high-order polynomial."

Similar to MAS-1, 5/18, Q. 45.

45. A time series, $\{x_t\}$ is modeled as an AR(3) process given by:

$$x_t = 4 + 0.45x_{t-1} + 0.25x_{t-2} + 0.05x_{t-3} + w_t$$

where w_t is white noise, with:

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

The most recent three observed values in this series are:

t	x_t
2016	30
2017	23
2018	21

Calculate the ten-step-ahead forecast value, x_{2028}

A. Less than 12
B. At least 12, but less than 16
C. At least 16, but less than 20
D. At least 20, but less than 24
E. At least 24

45. C. Since the expected value of w_t is zero, we can ignore it here.

$\hat{x}_{2019} = 4 + (0.45)(21) + (0.25)(23) + (0.05)(30) = 20.7.$

$\hat{x}_{2020} = 4 + (0.45)(20.7) + (0.25)(21) + (0.05)(23) = 19.715.$

$\hat{x}_{2021} = 4 + (0.45)(19.715) + (0.25)(20.7) + (0.05)(21) = 19.0968.$

Continuing in this manner, the forecasts are:
18.5573, 18.1107, 17.7440, 17.4403, 17.1897, 16.9826, **16.8116**.
Alternately, one can save some time, by letting y be the longterm limit.
Then: $y = 4 + 0.45y + 0.25y + 0.05y. \Rightarrow 0.25y = 4. \Rightarrow y = 16.$
Thus the forecast will decrease towards 16, and given the letter choices this is C.
Comment: One can rewrite the given AR(3) model as:
$(x_t - 16) = 0.45(x_{t-1} - 16) + 0.25(x_{t-2} - 16) + 0.05(x_{t-3} - 16) + w_t.$

Using a computer: $\hat{x}_{2068} = 16.000387.$

The variance of w_t would affect the variance and thus the accuracy of the forecasts.

The variance of $\hat{x}_{2019}$ is σ^2. The variance of $\hat{x}_{2020}$ is: $0.45^2\sigma^2 + \sigma^2 = 1.2025\sigma^2$.

The variance of $\hat{x}_{2021}$ is: $0.25^2\sigma^2 + (0.45^2)(1.2025\sigma^2) + \sigma^2 = 1.30600625\sigma^2$.

END OF EXAMINATION