Solutions to the
Fall 2018
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.

CAS Exam MAS-1

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. Stochastic Models, Section 3
2. Stochastic Models, Section 15
3. Stochastic Models, Section 11
4. Stochastic Models, Section 10
5. Life Contingencies, Section 1

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

11. Stochastic Models, Section 27
12. Life Contingencies, Section 4
13. Life Contingencies, Section 2
14. Simulation, Section 4
15. Statistics, Section 3

16. Statistics, Section 33
17. Statistics, Section 38
18. Statistics, Section 2
19. Statistics, Section 4
20. Statistics, Section 6

21. Statistics, Section 28
22. Statistics, Section 21
23. Statistics, Section 15
24. Loss and Frequency Distributions, Section 26
25. Statistics, Section 33

26. Generalized Linear Models, Section 15
27. Generalized Linear Models, Section 5
28. Generalized Linear Models, Section 16
29. Regression, Section 3
30. Regression, Section 10

31. Generalized Linear Models, Section 10
32. Regression, Section 10
33. Statistical Learning, Section 2
34. Regression, Section 13
35. Regression, Section 9

36. Statistical Learning, Section 14
37. Statistical Learning, Section 6
38. Statistical Learning, Section 7
39. Statistical Learning, Section 4
40. Statistical Learning, Section 12

41. Time Series, Section 3
42. Time Series, Section 16
43. Time Series, Section 9
44. Time Series, Section 16
45. Time Series, Section 3

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

	Number	Percent
Stochastic Models	7	15.6%
Loss & Freq. Dists.	1	2.2%
Statistics	10	22.2%
Regression	5	11.1%
GLMs	4	8.9%
Statistical Learning	6	13.3%
Life Contingencies	3	6.7%
Reliability	3	6.7%
Time Series	5	11.1%
Simulation	1	2.2%
Total	45	100.0%

1. Insurance claims are made according to a Poisson process $\{N(t), t \geq 0\}$ with rate $\lambda = 1$.
Calculate E[N(1) * N(2)].
A. Less than 1.5
B. At least 1.5, but less than 2.5
C. At least 2.5, but less than 3.5
D. At least 3.5, but less than 4.5
E. At least 4.5

1. C. N(2) - N(1) is independent of N(1).

$\Rightarrow$ E[{N(2) - N(1)} N(1)] = E[N(2) - N(1)] E[N(1)] = $\lambda\ \lambda = \lambda^2$.

$\Rightarrow$ E[N(2) N(1)] - E[N(1)^2] = λ^2.

N(1) is Poisson with mean λ and variance λ. $\Rightarrow$ E[N(1)^2] = $\lambda + \lambda^2$.

$\Rightarrow$ E[N(1) N(2)] = E[N(1)^2] + $\lambda^2 = \lambda + 2\lambda^2 = 1 + (2)(1^2) =$ **3**.
Comment: I waited to the end to substitute in the value of lambda, to illustrate the general ideas.

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

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

Calculate the probability that $T_2 > 3$.

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

2. A. $T_2 > 3$. $\Leftrightarrow$ 2nd event after time 3. $\Leftrightarrow$ 0 or 1 event in [0, 3].

Number of events by time 3 is N(3).

N(3) is Poisson with mean: $\int_0^3 \lambda(t)\, dt = \int_0^3 1.5t\, dt = 0.75t^2\Big]_{t=0}^{t=3} = (0.75)(3^2) = 6.75.$

$\text{Prob}[T_2 > 3] = f(0) + f(1) = e^{-6.75} + 6.75\, e^{-6.75} =$ **0.907%**.

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

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

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

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

3. E. Probability that a claim is from Portfolio 1 is: $\frac{\lambda_1}{\lambda_1 + \lambda_2}$ = 4/(4+2) = 2/3.

Prob[3 claims from Portfolio 1 before 3 claims from Portfolio 2] =
Prob[at least 3 of the first 5 claims from Portfolio 1].
The number of claims from Portfolio 1 out of the first 5 is Binomial with q = 2/3 and m = 5.

The desired probability is: f(3) + f(4) + f(5) = $\binom{5}{3}(2/3)^3(1/3)^2 + \binom{5}{4}(2/3)^4(1/3) + (2/3)^5$ =

(10)(8/243) + (5)(16/243) + 32/243 = **79.01%**.
Alternately, we want F(3-1) for a Negative Binomial Distribution wit parameters r = 3 and

$\beta = \frac{\lambda_2}{\lambda_1}$ = 2/4 = 1/2. F(2) = f(0) + f(1) + f(2) = $\frac{1}{(1+\beta)^r} + \frac{r\,\beta}{(1+\beta)^{r+1}} + \frac{r\,(r+1)\,\beta^2\ /2}{(1+\beta)^{r+2}}$ =

$1/1.5^3 + (3)(1/2)/1.5^4 + \{(3)(4)\ (1/2)^2 / 2\} / 1.5^5$ = **79.01%**.

Comment: If one has two independent Poisson Processes with claims intensities λ_1 and λ_2, then the chance that n claims from the first process occur before k claims from the second process is either the sum of Binomial terms from n to n + k - 1, with $q = \lambda_1 / (\lambda_1 + \lambda_2)$ and m = n + k - 1,
or F(k-1) for a Negative Binomial Distribution with parameters r = n and $\beta = \lambda_2 / \lambda_1$.

4. You are given the following information about a Poisson process:
- Claims occur at a rate of 4 per month
- Each claim is independent and takes on values with probabilities below:

Probability	Claim Amount
0.25	5
0.50	50
0.25	95

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

Calculate Var[X].

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

4. E. The second moment of severity is: $(0.25)(5^2) + (0.50)(50^2) + (0.25)(95^2) = 3512.5$.
Over twelve months, the average frequency is: $(4)(12) = 48$.
The variance of the Compound Poisson is: $(48)(3512.5) =$ **168,600**.
Comment: It is common to use S rather than X for aggregate losses.

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

$$S(x) = \left(1 - \frac{x}{30}\right)^{1/2}, \text{ for } 0 \le x < 30$$

A homeowner replaced his kitchen light with an LED light bulb 24 months ago and it is still functioning today.
Calculate the probability that the LED light bulb will stop working between 24 months and 81 months from today.

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

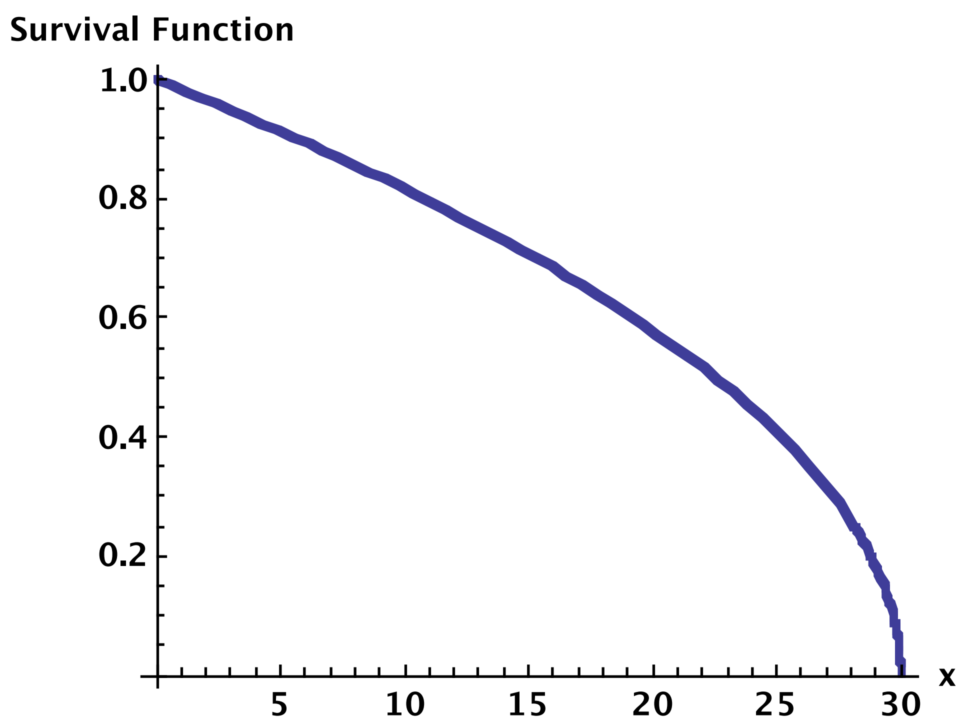
5. B. We want the conditional probability of failing between (24+24)/12 = 4 years and (24 + 81)/12 = 8.75 years, given survives 24/12 = 2 years.

$S(2) = (1 - 2/30)^{1/2} = 0.9661$. $S(4) = (1 - 4/30)^{1/2} = 0.9309$.

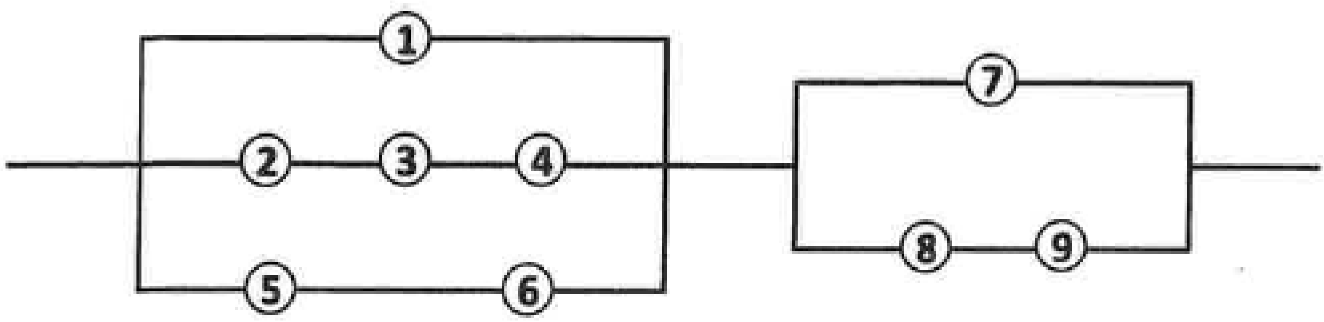
$S(8.75) = (1 - 8.75/30)^{1/2} = 0.8416$.

Prob[fails between 4 and 8.75 | survives until 2] = $\frac{S(4) - S(8.75)}{S(2)} = \frac{0.9309 - 0.8416}{0.9661}$ = **9.24%**.

<u>Comment</u>: While the survival function is given in terms of years, other quantities are given in months.
The survival function is an example of what is called Modified DeMoivre's Law.
A graph of this original survival function:



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



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

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

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

6. D. Since all of the components are otherwise identical, one can reason it out.
Components 1 and 7 are similarly placed. However, component 7 is on one of two parallel paths, while component 1 is on one of three parallel paths. Thus component 7 is more critical than component 1. Similarly component 8 is more critical than component 5.
Component 7 working means the second piece works, component 8 working means that the second piece works only if component 9 also works. Thus component 7 is more critical than component 8. Similarly component 1 is more critical than either components 5 or 2.
Putting it all together, **component 7** is most critical.
Alternately, one can compute the original and altered reliabilities.
The lower path of the second piece functions if both 8 and 9 function, this has probability: $p_8\ p_9$.
For the second piece, it fails if both paths fail; this has probability: $(1 - p_7)\ (1 - p_8\ p_9)$.
Similarly the probability that the first piece fails is: $(1 - p_1)\ (1 - p_2\ p_3\ p_4)\ (1 - p_5\ p_6)$.
Thus the reliability function of the whole system is:
$\{1 - (1 - p_7)\ (1 - p_8\ p_9)\}\{1 - (1 - p_1)\ (1 - p_2\ p_3\ p_4)\ (1 - p_5\ p_6)\}$.
Replacing Component 1, the reliability is 87.94%.
Replacing Component 2, the reliability is 84.53%.
Replacing Component 5, the reliability is 85.36%.
Replacing Component 7, the reliability is 91.63%.
Replacing Component 8, the reliability is 86.94%.
Thus from least improvement to most improvement in reliability:
Component 2 < Component 5 < Component 8 < Component 1 < **Component 7**.
Comment: You should be able to determine the reliability function. However, calculating the reliability for all five possibilities is tedious and time consuming under exam conditions.
For the original system with all of the $p_i = 0.75$, the reliability is 83.43%.

7. A 3-out-of-50 system is placed in series with a 48-out-of-50-system.
Calculate the number of minimal path sets.
A. Fewer than 20,000
B. At least 20,000 but fewer than 100,000
C. At least 200,000 but fewer than 2,000,000
D. At least 2,000,000 but fewer than 20,000,000
E. At least 20,000,000

7. E. A minimal path set is such that the system functions but would not function if any element of the minimal path fails.
3 out of 50 system functions if and only if at least 3 components function.
Thus we choose any 3 components and they will be minimal path set.

The number of such paths is: $\binom{50}{3}$ = 19,600.

Similarly, the number of minimal path sets for the 48-out-of-50-system is:

$\binom{50}{48}$ = 1225.

The series system fails if any of the two subsystem fails.
Thus a minimal path set would consist of any minimal path set of the 3-out-of-50 system together with any minimal path set of the 48-out-of-50 system. Thus the number of minimal path sets for the system is: (19,600)(1225) = **24,010,000**.
Comment: Choice B was intended to be: At least 20,000 but fewer than **2**00,000.
If instead the two subsystems were placed in parallel, then the number of minimal path sets would be: 19,600 + 1225 = 20,825. See CAS S, 11/17, Q.8.

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

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

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

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

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

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

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

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

8. A. For the uniform, $F(t) = t$, for $0 \le t \le 1$.

For the Exponential, $F(t) = 1 - e^{-t/2}$, for $t > 0$.

The parallel system fails if both components fail.

Thus $S(t) = 1 - t\,(1 - e^{-t/2}) = 1 - t + t\, e^{-t/2}$, for $0 \le t \le 1$.

For $t \ge 1$, the uniform component has failed, so the survival function for the system is that of the Exponential: $S(t) = e^{-t/2}$, $t > 1$.

Expected lifetime = $\int S(t)\, dt = \int_0^1 1 - t + t\, e^{-t/2}\, dt + \int_1^\infty e^{-t/2}\, dt$.

<u>Comment</u>: One can eliminate some of the choices.

Choice C involves an integral from 1 to infinity that is infinite.

Choice B has a single integral from 0 to infinity, which can not be right due to the change in the form of Survival Function at 1.

Doing the two integrals, the expected lifetime is: 0.861 + 1.213 = 2.074.

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

- There are three states to answering a trivia question: Skip (State 0), Correct (State 1), and Wrong (State 2).
- $P = \begin{pmatrix} 0 & 0.85 & 0.15 \\ 0.20 & 0.80 & 0 \\ 0 & 0.70 & 0.30 \end{pmatrix}$.
- The candidate skipped the previous question.

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

A. Less than 0.92

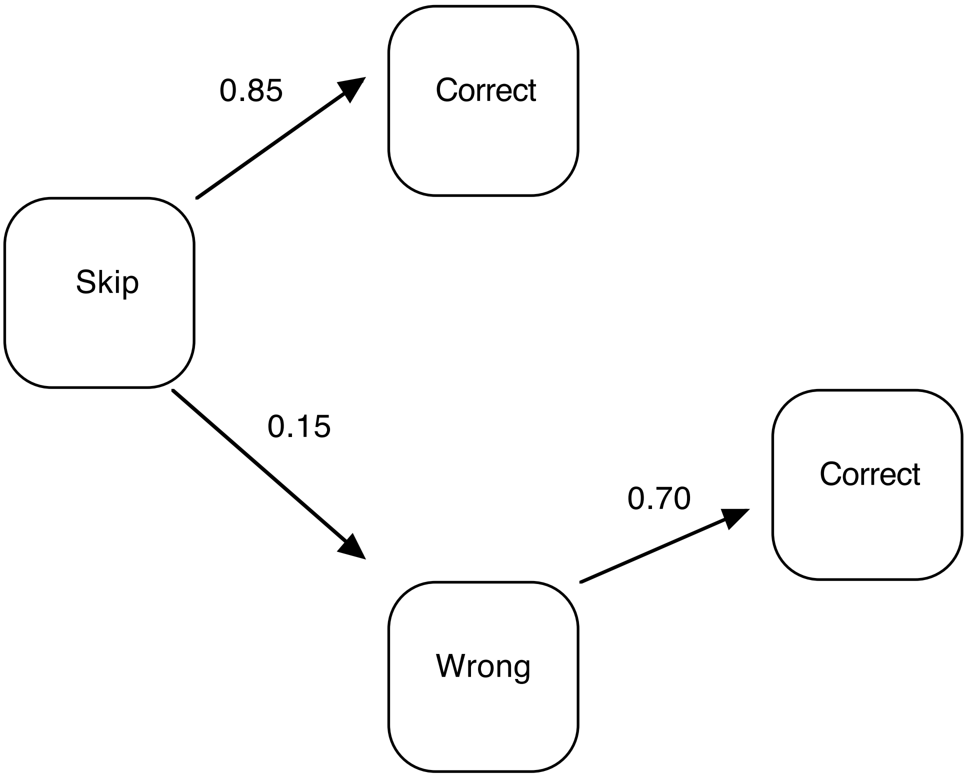
B. At least 0.92, but less than 0.94

C. At least 0.94, but less than 0.96

D. At least 0.96, but less than 0.98

E. At least 0.98

9. C. Draw a branch diagram (where I have left out branches not needed to answer the question):



The probability of the candidate correctly answering at least one of the two subsequent questions:
0.85 + (0.15)(0.70) = **0.955**.
Alternately, (1, 0, 0)P = (0, 0.85, 0.15). (0, 0.85, 0.15) P = (0.17, 0.785, 0.045).
Thus the chance of being in State 1 (correct) at time 1 plus the chance of being in State 1 (correct) at time 2 is: 0.85 + 0.785 = 1.635. But this includes the chance of being in State 1 at both times 1 and 2; this probability is (0.85)(0.8) = 0.68.
Thus the answer to question asked is: 1.635 - 0.68 = **0.955**.

10. You are given the following information about when a fair coin is flipped:
- If the outcome is Heads, 1 chip is won
- If the outcome is Tails, I chip is lost
- A gambler starts with 20 chips and will stop playing when he either has lost all his chips or he reaches 50 chips
- Of the first 10 flips, 7 are Heads and 3 are Tails

Calculate the probability that the gambler will lose all of his chips, given the results of the first 10 flips.
A. Less than 0.5
B. At least 0.5, but less than 0.6
C. At least 0.6, but less than 0.7
D. At least 0.7, but less than 0.8
E. At least 0.8

10. B. After the first 7 coin flips, the gambler's chips: 20 + 7 - 3 = 24.
Gambler's Ruin Problem, with p = 1/2, i = 24, and N = 50.
The chance that the Gambler ends up with 50 chips is: i/N = 24/50 = 48%.
Thus the chance that the Gambler loses all of his chips is: 1 - 48% = **52%.**

11. A scientist has discovered a way to create a new element.
This scientist studied the new element and observed some of its properties:
- The element has a lifespan of 1 week, then it evaporates
- The element can produce offspring
- The creation of offspring is distributed as follows:

Number of Offspring	Probability
0	0.75
1	0.20
2	0.05

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

Determine which of the following best describes the above process.
A. Branching process
B. Gambler's ruin
C. Periodic Markov chain
D. Time reversible Markov chains
E. None of (A), (B), (C), or (D) describe this process

11. A. This is an example of a branching process.
Comment: $\mu = (0.75)(0) + (0.20)(1) + (0.05)(2) = 0.30 \le 1. \Rightarrow \pi_0 = 1.$

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

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

Year, t	p_t
0	0.90
1	0.80
2	0.70

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

Calculate the actuarial present value of this policy.

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

12. C. (0.90)(0.80) = 0.72. (0.72)(0.70) = 0.504.

Year, t	p_t	S(t+1)	Difference
0	0.90	0.90	0.10
1	0.80	0.72	0.18
2	0.70	0.504	0.216

1 - 0.9 = 0.1. 0.9 - 0.72 = 0.18. 0.72 - 0.504 = 0.216.

APV = $(0.1)(500)/1.05 + (0.18)(500)/1.05^2 + (0.216)(500)/1.05^3 + (0.504)(100)/1.05^3$ = **266**.

13. You are given the following information:

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

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

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

13. C. ${}_{55}p_{35} = l_{90} / l_{35}$ = 1,058,491 / 9,420,657 = 0.1124.

${}_{25}p_{55} = l_{80} / l_{55}$ = 3,914,365 / 8,640,861 = 0.4530.

Prob[(55) alive at age 90] = (0.4530) $(1 - 0.0803)^{10}$ = 0.1961.

Prob[(35) alive at age 90] Prob[(55) not alive at age 90] = (0.1124)(1 - 0.1961) = **9.04%**.

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

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

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

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

14. E. $F(x) = \int_0^x f(t)\,dt = x^5$. Set $0.6 = F(x) = x^5$. $\Rightarrow$ $x =$ **0.9029**.

<u>Comment</u>: A Beta Distribution with $a = 5$, $b = 1$, and $\theta = 1$.

15. You are given the following:

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

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

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

A. $\frac{4\theta^2}{3n}$ B. $\frac{\theta^2}{3n^2}$ C. $\frac{\theta^2}{12}$ D. $\frac{\theta^2}{6}$ E. $\frac{\theta^2}{3n}$

15. E. $\hat{\theta} = 2\bar{X}$. $\text{Var}[\hat{\theta}] = 2^2 \text{Var}[\bar{X}] = 4 \text{Var}[X]/n = (4)(\theta^2/12)/n = \boldsymbol{\theta^2/(3n)}$.

Alternately, $\text{Var}[\hat{\theta}] = (2/n)^2 \sum_{i=1}^{n} \text{Var}[X_i] = (4/n^2)\ n\ \text{Var}[X] = (4/n)(\theta^2/12) = \boldsymbol{\theta^2/(3n)}$.

<u>Comment</u>: In general, variances of estimated parameters go down as 1/n,
eliminating choices B, C, and D.
Since the mean of this uniform distribution is $\theta / 2$, θ is being estimated via the method of moments.
The variance of a uniform distribution on (a, b) is: $(b-a)^2/12$.

16. Losses occur independently with the probabilities as described below:

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

Using a sample of size 2, the variance of individual losses is estimated using the sample variance given by: $\sum_{i=1}^{2} (x_i - \bar{x})^2$

Calculate the mean square error of this estimator.

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

16. D. E[X] = (25%)(0) + (60%)(2) + (15%)(6) = 2.1.

$E[X^2]$ = $(25\%)(0^2) + (60\%)(2^2) + (15\%)(6^2) = 7.8$.

Var[X] = $7.8 - 2.1^2 = 3.39$.

List all possible samples of size 2, the computed sample variance, and the squared error.

First	Second	Sample Var.	Sq. Error	Prob.	Product
0	0	0	11.4921	6.25%	0.718256
0	2	2	1.9321	15.00%	0.289815
0	6	18	213.4521	3.75%	8.004454
2	0	2	1.9321	15.00%	0.289815
2	2	0	11.4921	36.00%	4.137156
2	6	8	21.2521	9.00%	1.912689
6	0	18	213.4521	3.75%	8.004454
6	2	8	21.2521	9.00%	1.912689
6	6	0	11.4921	2.25%	0.258572
				1	**25.527900**

For example, {0, 2} has probability: (25%)(60%) = 15%.

The mean is 1 and the sample variance is: $(0 - 1)^2 + (2 - 1)^2 = 2$.

The squared error is: $(2 - 3.39)^2 = 1.9321$. (15%)(1.9321) = 0.289815.

Comment: One can save some time by adding up the probabilities of those cases with the same value of the sample variance.

In this case with only a sample size of two, the sample variance is: $(x_1 - x_2)^2/2$.

The expected value of the sample variance is:
(6.25%)(0) + (2)(15%)(2) + (2)(3.75%)(18) + (36%)(0) + (2)(9%)(8) + (2.25%)(0) = 3.39.
In general, the sample variance is an unbiased estimator of the variance; therefore its mean squared error is equal to its variance.

In general, the sample variance has a mean squared error of:

$E[(X - E[X])^4] / n + (3 - n)E[(X - E[X])^2]^2 / (n^2 - n)$.

In this case, $n = 2$, $E[(X - E[X])^2] = Var[X] = 3.39$, *and*

$E[(X - E[X])^4] = (25\%)(2.1^4) + (60\%)(0.1^4) + (15\%)(3.9^4) = 39.5637$.

$E[(X - E[X])^4] / n + (3 - n)E[(X - E[X])^2]^2 / (n^2 - n) = 39.5637/2 + 3.39^2/2 = 25.5279$.

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

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

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

A. Less than 30

B. At least 30, but less than 50

C. At least 50, but less than 70

D. At least 70, but less than 90

E. At least 90

17. C. For a Normal Distribution with mean 0, $f(x) = \frac{\exp[-\frac{x^2}{2\sigma^2}]}{\sigma\sqrt{2\pi}}$.

The loglikelihood is: $\frac{-1}{2\sigma^2} \sum_{i=1}^{n} X_i^2$ - n ln(σ) - (n/2) ln(2π).

Thus, by Neyman's Theorem, $\sum_{i=1}^{n} X_i^2$ is a sufficient statistic.

By the Rao-Blackwell theorem, if a minimum variance unbiased estimator of σ^2 exists, then it is a function of the sufficient statistic $\sum_{i=1}^{n} X_i^2$.

$\sum_{i=1}^{n} X_i^2$ / n is an unbiased estimate of the variance of a distribution with mean zero.

This Normal Distribution is a (one parameter) exponential family.

Thus $\sum_{i=1}^{n} X_i^2$ / n is an MVUE, which takes on the value: 500/10 = **50**.

Alternately, If Y is a sufficient statistic for a parameter, then the maximum likelihood fit will be function of Y. For the Normal Distribution with known mean μ, the maximum likelihood estimator of σ is:

$\sqrt{\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - 0)^2}{n}}$. Thus $\sum_{i=1}^{n} X_i^2$ / n is a sufficient statistic. Proceed as before.

Comment: Difficult!

For a distribution with known mean, the usual and unbiased estimator of the variance is: $\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}$.

The sample variance is used when we have an unknown mean.

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

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

Losses are assumed to follow a uniform distribution within each interval.
Calculate the fitted probability of a loss exceeding 500.

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

18. C. We have assumed that the average size of loss in each interval is its midpoint.
Set the observed mean equal to the mean of the Exponential Distribution.

$$\theta = \frac{(5000)(50) + (3000)(300) + (1500)(750) + (500)(3000)}{5000 + 3000 + 1500 + 500} = 377.5.$$

$S(500) = e^{-500/377.5} =$ **26.6%**.

Comment: It would have been better to say that the average size of loss in each interval is assumed to be at its midpoint, rather than "Losses are assumed to follow a uniform distribution within each interval."

19. An actuary observes the following 20 losses for a select insurance policy:
{20, 25, 36, 38, 42, 52, 55, 57, 65, 66, 69, 71, 72, 73, 74, 74, 74, 80, 81, 82}
She believes the distribution which fits the data the best is loglogistic with the following probability density:

$$f(x;\ \gamma,\ \theta) = \frac{\gamma\, x^{\gamma-1}}{\theta^{\gamma}(1\ +\ (x/\theta)^{\gamma})^2}.$$

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

Calculate the estimated value of θ.

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

19. E. I estimate the 20th percentile as the (20+1)(0.2) = 4.2th loss from smallest to largest:
(0.8)(38) + (0.2)(42) = 38.8.

I estimate the 80th percentile as the (20+1)(0.8) = 16.8th loss from smallest to largest: 74.

From the tables attached to the exam, for the loglogistic: $F(x) = \frac{(x/\theta)^{\gamma}}{1\ +\ (x/\theta)^{\gamma}}$.

Matching at the two percentiles:

$$0.2 = \frac{(38.8/\theta)^{\gamma}}{1 + (38.8/\theta)^{\gamma}}. \Rightarrow 5 = \frac{1 + (38.8/\theta)^{\gamma}}{(38.8/\theta)^{\gamma}} = (\theta/38.8)^{\gamma} + 1. \Rightarrow (\theta/38.8)^{\gamma} = 4.$$

$$0.8 = \frac{(74/\theta)^{\gamma}}{1 + (74/\theta)^{\gamma}}. \Rightarrow 1.25 = \frac{1 + (74/\theta)^{\gamma}}{(74/\theta)^{\gamma}} = (\theta/74)^{\gamma} + 1. \Rightarrow (\theta/74)^{\gamma} = 0.25.$$

Dividing the two equations: $(74/38.8)^{\gamma} = 16. \Rightarrow \gamma = \ln 16 / \ln(74/38.8) = 4.294.$

Thus $(\theta/38.8)^{4.294} = 4. \Rightarrow 4^{(1/4.294)}\ (38.8) = \mathbf{53.6}$.

<u>Comment</u>: Check. $\frac{(38.8/53.6)^{4.294}}{1 + (38.8/53.6)^{4.294}} = 0.2.$ $\frac{(74/53.6)^{4.294}}{1 + (74/53.6)^{4.294}} = 0.8.$

How to estimate percentiles is not on the syllabus. I have used the smoothed empirical estimate of percentiles, as per Equations 11.7 to 11.10 in
<u>NonLife Actuarial Model Theory, Methods and Evaluation</u>, not on the syllabus.

If one instead estimated the 20th percentile as 38, then: $\gamma = \ln 16 / \ln(74/38) = 4.160$,

and thus $\theta = 4^{(1/4.160)}\ (38) = 53.0$.

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

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

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

Determine which of the following is the MLE for θ.

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

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

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

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

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

20. D. For the Exponential (with ungrouped data) the maximum likelihood fit is the sum of the payments divided by the number of uncensored values: $\dfrac{\sum_{i=1}^{n} X_i + su}{n} = \boldsymbol{\frac{1}{n} \sum_{i=1}^{n} X_i + \frac{su}{n}}$.

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

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

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

You want to perform the following hypothesis test:

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

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

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

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

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

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

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

21. B. Loglikelihood is: $n \ln(\lambda) + (\lambda - 1) \sum_{i=1}^{n} \ln X_i + m \ln(\mu) + (\mu - 1) \sum_{i=1}^{m} \ln Y_i$.

The best critical region has the form: Loglikelihood for H_1 - Loglikelihood for $H_0 \ge k$.

$$n \ln(3) + 2 \sum_{i=1}^{n} \ln X_i + m \ln(8) + 7 \sum_{i=1}^{m} \ln Y_i - \{n \ln(2) + \sum_{i=1}^{n} \ln X_i + m \ln(4) + 3 \sum_{i=1}^{m} \ln Y_i\} \ge k.$$

$$\sum_{i=1}^{n} \ln X_i + 4 \sum_{i=1}^{m} \ln Y_i \ge k - n\{\ln(3) - \ln(2)\} - m\{\ln(8) - \ln(4)\} = c.$$

<u>Comment</u>: f and g are each Beta Distributions with b = 1 and $\theta = 1$.

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

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

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

Calculate the p-value of this test.

A. Less than 0.01

B. At least 0.01, but less than 0.02

C. At least 0.02, but less than 0.05

D. At least 0.05, but less than 0.10

E. At least 0.10

22. E. Work with the four differences: 3, -1, 7, and 3.

Their mean is 3 and sample variance is 10.667.

$t = \frac{3 - 0}{\sqrt{10.667/4}} = 1.837$, with 4 - 1 = 3 degrees of freedom.

Performing a two-sided test, consulting the tables: 1.638 < 1.837 < 2.353.

Thus the p-value is between 20% and 10%.

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

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

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

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

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

23. B. Let X = 30 - c and Y = 30 + c, for c > 0.
(Then X < Y and the total number of rolls adds to the given 180.)
The expected number for each result is 180/6 = 30.

The Chi-Square Statistic is: $\frac{(\text{observed number - expected number})^2}{\text{expected number}} =$

$0^2/30 + 0^2/30 + 0^2/30 + 0^2/30 + c^2/30 + c^2/30 = c^2/15$.

For 6 - 1 = 5 degrees of freedom, looking in the table the 5% critical value is 11.07.

Set $11.07 = c^2/15. \Rightarrow c = 12.89$.

For c = 13 and thus X = **17**, the Chi-Square Statistic is $13^2/15 = 11.27 > 11.07$; we reject H_0.

For c = 12 and X = 18, the Chi-Square Statistic is $12^2/15 = 9.60 < 11.07$; we do not reject H_0.

24. The distribution of Y is given as:

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

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

Calculate the mean of X.

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

24. D. $X > x. \Leftrightarrow 5(e^y - 1) > x. \Leftrightarrow e^y > 1 + x/5. \Leftrightarrow y > \ln(1 + x/5)$.

$S(x) = \text{Prob}[X > x] = \text{Prob}[y > \ln(1 + x/5)] = \exp[-11(\ln(1 + x/5)] = (1 + x/5)^{-11} = \left(\frac{5}{5 + x}\right)^{11}$.

This is a Pareto Distribution with $\alpha = 11$ and $\theta = 5$. The mean is: $\theta/(\alpha - 1) = 5/(11-1) =$ **0.5**.

Alternately, $f(y) = 11\ e^{-11y}$.

$\frac{dx}{dy} = 5e^y$.

Thus the density of x is: $11\ e^{-11y}\ /\ (5e^y) = (11/5)\ e^{-12y}$.

However, $e^y = 1 + x/5$.

Thus the density of x is: $(11/5)\ (1 + x/5)^{-12} = (11)\ (5^{11})\ /\ (5 + x)^{12}$.

This is the density of a Pareto Distribution with $\alpha = 11$ and $\theta = 5$.

The mean is: $\theta/(\alpha - 1) = 5/(11-1) =$ **0.5**.

Alternately, $E[X] = \int_0^\infty X\ f(y)\ dy = \int_0^\infty (5e^y - 1)\ (11\ e^{-11y})\ dy = 55 \int_0^\infty e^{-10y} - e^{-11y}\ dy$

$= 55(1/10 - 1/11) =$ **0.5**.

<u>Comment</u>: In general, if $F(y) = 1 - e^{-\lambda y}$ and $X = \theta(e^Y - 1)$, then X follows a Pareto distribution with parameters $\alpha = \lambda$ and θ.

The form of letter choice A is unusual; it would usually read “Less than 0.35”.

25. You draw a large number of independent samples, each of size n = 4, from a uniform distribution on (0, θ).
You want to use the second smallest value in each sample as an estimate for the mean.
The density for the k^{th} order statistic of a sample is given as:

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

Calculate the expected bias of this estimate.

A $-\frac{4\theta}{5}$ B. $-\frac{4\theta}{7}$ C. $-\frac{\theta}{10}$ D. $-\frac{\theta}{30}$

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

25. C. $E[Y] = \theta/2$. $E[Y_{(2)}] = \theta\, 2/(4+1) = 2\theta/5$. Bias $= 2\theta/5 - \theta/2 =$ $\mathbf{-\frac{\theta}{10}}$.

Comment: For a uniform distribution on (0, θ), the expected values of the order statistics are:

$\frac{\theta}{n+1}\ \{1, 2, 3, ... , n\}$.

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

Response variable	retention
Response distribution	binomial
Link	probit
Pseudo R^2	0.6521

Parameter	df	$\hat{\beta}$
Intercept	1	0.4270
Tenure		
< 5 years	0	0.0000
≥ 5 years	1	0.1320
Prior Rate Change		
< 0%	1	0.0160
[0%,10%]	0	0.0000
> 10%	1	-0.0920
Amount of Insurance (000's)	1	0.0015

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

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

26. C. $X\beta$ = 0.4270 + 0 - 00920 + (225)(0.0015) = 0.6725.

$\hat{\pi}$ = Φ [0.67] = **0.7486**.

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

I. $g(y) = 2 + 3y + 3(y - 5)^2$

II. $g(y) = 4 - 4y$

III. $g(y) = |y|$

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

A. I only

B. II only

C. III only

D. I, II and III

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

27. B. The link function, g, has to be differentiable and monotonic.

I. This function is differentiable. $g'(x) = 3 + 6(y - 5)$.

Setting the derivative equal to zero: $0 = 3 + 6(y - 5)$. $\Rightarrow$ $y = 4.5$.

g reaches a minimum at y = 4.5. It deceases for y < 4.5 and increase for y > 4.5; it is not monotonic.

II. This function is monotonic, always decreasing. This function is differentiable.

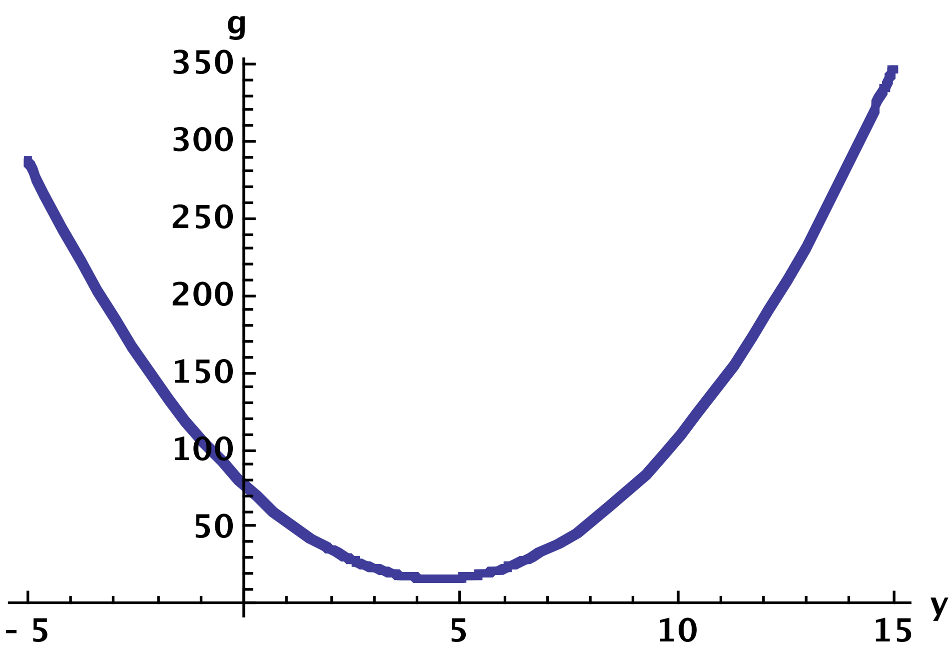
III. This function is decreasing for y < 0 and increasing for y > 0; it is not monotonic.

(It is differentiable except at y = 0.)

<u>Comment</u>: See page 51 of Dobson and Barnett. A function is monotonic if and only if it is either entirely non-increasing, or entirely non-decreasing. However, if g is flat for some part of its domain, then it is not one-to-one and one can not invert g. Thus I think we really want g to be strictly increasing or strictly decreasing, so that the function is one-to-one, in other words strictly monotonic.

For the second function: $g^{-1}(x) = 1 - x/4$.

A graph of the first function:



28. In a study 100 subjects were asked to choose one of three election candidates (A, B or C).
The subjects were organized into four age categories: (18-30, 31-45, 45-61, 61+).
A logistic regression was fitted to the subject responses to predict their preferred candidate, with age group (18-30) and Candidate A as the reference categories.
For age group (18-30), the log-odds for preference of Candidate B and Candidate C were -0.535 and -1.489 respectively.
Calculate the modeled probability of someone from age group (18-30) preferring Candidate B.
A. Less than 20%
B. At least 20%, but less than 40%
C. At least 40%, but less than 60%
D. At least 60%, but less than 80%
E. At least 80%

28. B. Since Candidate A is the base level: $\hat{\pi}_B = \hat{\pi}_A \exp[X\beta]$. $\hat{\pi}_C = \hat{\pi}_A \exp[X\beta]$.

The log of the odds ratios are: $\ln[\hat{\pi}_B / \hat{\pi}_A]$ and $\ln[\hat{\pi}_C / \hat{\pi}_A]$.

$\Rightarrow$ $-0.535 = \ln[\hat{\pi}_B / \hat{\pi}_A]$. $\Rightarrow$ $\hat{\pi}_B = \hat{\pi}_A e^{-0.535} = 0.58567\ \hat{\pi}_A$.

$-1.489 = \ln[\hat{\pi}_C / \hat{\pi}_A]$. $\Rightarrow$ $\hat{\pi}_C = \hat{\pi}_A e^{-1.489} = 0.22560\ \hat{\pi}_A$.

Since the probability of being in the three categories must add to one:

$1 = \hat{\pi}_A + 0.58567\ \hat{\pi}_A + 0.22560\ \hat{\pi}_A$. $\Rightarrow$ $\hat{\pi}_A = 0.55210$.

$\Rightarrow$ $\hat{\pi}_B = (0.58567)(0.55210) = \mathbf{0.32335}$. $\hat{\pi}_C = (0.22560)(0.55210) = 0.12455$.

<u>Comment</u>: An example of a GLM with a nominal response variable, a categorical variable without a natural order.

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

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

Calculate the residual for the 3rd observation.

A. Less than -2

B. At least -2, but less than 0

C. At least 0, but less than 2

D. At least 2, but less than 4

E. At least 4

29. B. $\hat{\beta} = \dfrac{N \sum X_i Y_i - \sum X_i \sum Y_i}{N \sum X_i^2 - \left(\sum X_i\right)^2} = \dfrac{(5)(3263.9) - (29.4)(554)}{(5)(173.14) - 29.4^2} = 23.806.$

$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X} = (554/5) - (23.806)(29.4/5) = -29.179.$

$\hat{Y}_3 = -29.179 + (23.806)(6.0) = 113.657.$

Residual = actual - estimated.

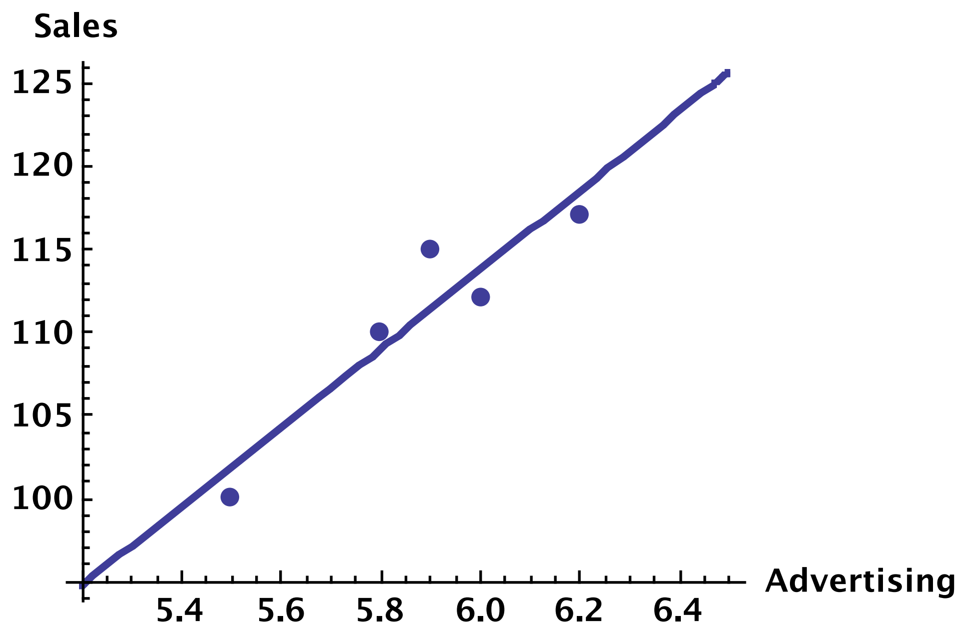
Thus the third residual is: 112 - 113.657 = **-1.657**.

<u>Comment</u>: The fitted responses are: 101.754, 108.896, 113.657, 111.276, 118.418.

The residuals are: -1.754, 1.104, -1.655, 3.724, -1.418.

You should know how to fit a linear regression using the stat function of your electronic calculator.

A plot of the data and the fitted line:



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

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

Calculate the adjusted R^2 for this model.

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

30. C. Let k = number of independent variables excluding the intercept. k = 12.
F has degrees of freedom of k and N - k - 1.
Thus 493 = N - k - 1 = N - 13. ⇒ N = 506.
$1 - R_a^2 = (1 - R^2)(N - 1) / (N - k - 1) = (1 - 0.7406)(505/493) = 0.2657. \Rightarrow R_a^2 =$ **0.7343**.

Alternately, $\hat{\sigma}^2 = \dfrac{\text{Residual Sum of Squares}}{N - k - 1}$, with N - k - 1 degrees of freedom.

⇒ 493 = N - k - 1 = N - 13. ⇒ N = 506.
Also we are given: $\hat{\sigma} = 4.74$.
⇒ 4.74^2 = (Residual Sum of Squares) / 493. ⇒ Residual SS = $(493)(4.74^2)$ = 11,077.
R^2 = 1 - (Residual Sum of Squares) / (Total Sum of Squares).
⇒ 0.7406 = 1 - 11,077 / TSS. ⇒ TSS = 42,702.

$R_a^2 = 1 - \dfrac{\text{(Residual Sum of Squares) / (N-k-1)}}{\text{(Total Sum of Squares) / (N-1)}}$ = 1 - (11,077 / 493) / (42,702 / 505) = **0.7343**.

Comment: The adjusted R^2 is always less than R^2, eliminating choices D and E.

I suspect that it was intended to say that the p-value is less than 2.2×10^{-16}.
There is extra information given in this question, that lets one solve in multiple ways.
Model SS = 42,702 - 11,077 = 31,625.

$F = \dfrac{\text{Model SS} / k}{\text{Residual SS} / (N - k - 1)}$ = (31,625/12) / (11,077/493) = 117.3, matching the given value.

$F = \dfrac{N - k - 1}{k} \dfrac{R^2}{1 - R^2} = \dfrac{493}{12} \dfrac{0.7406}{1 - 0.7406}$ = 117.3, matching the given value.

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

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

The actuary wants to evaluate which of the two models is superior.
Determine which of the following is the best course of action for the actuary to take.
A. Perform a likelihood ratio test
B. Compute the F-statistic and perform an F-test
C. Compute and compare the deviances of the two models
D. Compute and compare the AIC statistics of the two models
E. Compute the Chi-squared statistic and perform a Chi-squared test

31. D. One can not use the likelihood ratio test, since neither model is a special case of the other; the models are not nested.
One could only use the F-test if we had a multiple regression. Even then the F-Test could be used to test whether groups of slopes are significantly different form zero; which may not help to choose between the two models.
One should not just compare the deviances without taking into account the number of fitted parameters (number of degrees of freedom used) by each model.
The Chi-squared test does not apply to this situation of comparing two models.
Comment: Model 2 has both a better (smaller) loglikelihood and uses fewer degrees of freedom (is simpler) than Model 1; thus we prefer Model 2 to Model 1, with no need to do any more.
AIC = (-2) (maximum loglikelihood) + (number of parameters)(2).
For Model 1, AIC = (-2)(-11,565) + (10)(2) = 23,150.
For Model 2, AIC = (-2)(-11,562) + (8)(2) = 23,140.
Since it has the smaller AIC, we prefer Model 2.

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

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

	Sum of Squares
Regression	2.65376
Total	7.62956

He wants to test the following hypothesis:
- H_0: $\beta_1 = \beta_2 = \beta_3 = 0$
- H_1: At least one of $\{\beta_1, \beta_2, \beta_3\} \neq 0$

Calculate the value of the F-statistics used in this test.
A. Less than 1
B. At least 1, but less than 3
C. At least 3, but less than 5
D. At least 5
E. The answer cannot be computed from the information given.

32. B. Error (Residual) SS = Total SS - Regression (Model) SS = 7.62956 - 2.65376 = 4.9758.
In order to test whether all of the slopes are zero:

$F = \dfrac{\text{Regression SS} / k}{\text{Error SS} / (N - k - 1)}$ = (2.65376/3) / (4.9758/16) = **2.844**, with 3 and 16 degrees of freedom.

Comment: Using a computer, the p-value is 7.1%.
Due to the large absolute values of the t-statistics for each of the slopes, individually each slope is significantly different from zero. Each t-statistic has 20 - 3 - 1 = 16 degrees of freedom;
the p-values are: 0.9%, 0.2%, and 0.8%. Nevertheless, it may turn out that the F-test of the all of the slopes together gives a somewhat different result.

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

Model 1		
Parameter	$\hat{\beta}$	p-value
Intercept	212	
Replacement Cost (000s)	0.03	< .001
Roof Size	0.15	< .001
Precipitation Index	120	0.02
Replacement Cost (000s) x Roof Size	0.0010	0.05

Model 2		
Parameter	$\hat{\beta}$	p-value
Intercept	315	
Replacement Cost (000s)	0.02	< .001
Roof Size	0.17	< .001
Number of Bathrooms	210	0.03
Replacement Cost (000s) x Roof Size	0.0015	< .001

Model 1 Statistics		Model 2 Statistics	
R^2	0.91	R^2	0.94
Adj R^2	0.87	Adj R^2	0.89
MSE	31,765	MSE	30,689
AIC	25,031	AIC	25,636

Model 1		Model 2	
Cross Validation Set	MSE	Cross Validation Set	MSE
1	33,415	1	26,666
2	38,741	2	38,554
3	32,112	3	39,662
4	37,210	4	36,756
5	29,501	5	30,303

You use 5-fold cross validation to select superior of the two models.
Calculate the predicted expected annual loss for a homeowners policy with a 500,000 replacement cost, a 2,000 roof size, a 0.89 precipitation index and three bathrooms, using the selected model.
A. Less than 1,000
B. At least 1,000, but less than 1,500
C. At least 1,500, but less than 2,000
D. At least 2,000, but less than 2,500
E. At least 2,500

33. C. The estimated test MSE for Model 1 is:
(33,415 + 38,741 + 32,112 + 37,210 + 29,501) / 5 = 34,195.8.
The estimated test MSE for Model 2 is:
(26,666 + 38,554 + 39,662 + 36,756 + 30,303) / 5 = 34,388.2.
By this criterion, Model 1 with the smaller test MSE is preferred.
For Model 1, expected annual loss for the selected insured is:
212 + (0.03)(500) + (0.15)(2000) + (120)(0.89) + (0.0010)(500)(2000) = **1633.8**.
Comment: For Model 2, expected annual loss for the selected insured is:
315 + (0.02)(500) + (0.17)(2000) + (210)(3) + (0.0015)(500)(2000) = 2795.
Thus even if one does not know how to select which model is better, the choices are C or E.
The difference in test MSE between the two models is very small, particularly given the fluctuations in MSE among the different cross-validation sets.
Model 1 has the smaller (better) AIC, so it would be preferred using that criterion.
Model 2 has the larger (better) adjusted R^2, so it would be preferred using that criterion.
We should not use either R^2 or training MSE to compare the two models.
The two models use the same number of parameters. The two models differ in that Model 1 uses Precipitation Index, while Model 2 instead uses Number of Bathrooms.
In general, an actuary would sometimes select a simpler model even if a more complex model has a somewhat smaller test MSE.

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

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

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

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

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

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

34. D. $\text{VIF}_3 = \frac{\text{Var}[\hat{\beta}_3] \text{ for the Full Model}}{\text{Var}[\hat{\beta}_3] \text{ for the Model with just } \hat{\beta}_3 \text{ (plus the intercept)}} = \frac{0.80^2}{0.75^2} = \mathbf{1.138}$.

Comment: See page 101 of An Introduction to Statistical Learning.
I believe that this formula in the textbook is not correct.

35. You are fitting a linear regression model of the form:

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

and are given the following values used in this model:

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

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

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

Calculate the residual for the 5^{th} observation.

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

35. A. $\hat{\beta} = (X^T X)^{-1} X^T y$.

Thus the fitted coefficients are: 2.335, 0.297, -0.196, and 1.968.
The X values for the 5th observation are on the fifth row of the design matrix of X.

$\hat{Y}_5 = 2.335 + (0.297)(1) + (-0.196)(0) + (1.968)(6) = 14.44.$

Fifth residual = $Y_5 - \hat{Y}_5 = 13 - 14.44 =$ **-1.44**.

Alternately, $\hat{Y} = Hy$. Thus we can get $\hat{Y}_5$ by multiplying the fifth row of H by the given y vector.

$\hat{Y}_5 = (-0.146)(19) + (-0.038)(32) + (0.184)(19) + (0.411)(17) + (0.443)(13) + (0.146)(15) =$

14.44. Fifth residual = $Y_5 - \hat{Y}_5 = 13 - 14.44 =$ **-1.44**.

Comment: This is a model with an intercept, since the design matrix for X has ones in the first column.
There are 3 independent variables, based on the remaining three columns of the design matrix.
There are six observed points, based on the six rows of the design matrix for X.
Using a computer to fit the model, the six residuals are: -0.854, 0.038, 0.816, 0.589, -1.443, 0.854.
H is the Hat Matrix. The diagonal elements of H are the leverage values.
The average leverage is (k+1)/N = 4/6 = 0.667; in this case none of the points has a high leverage (more than 2 or 3 times the average).

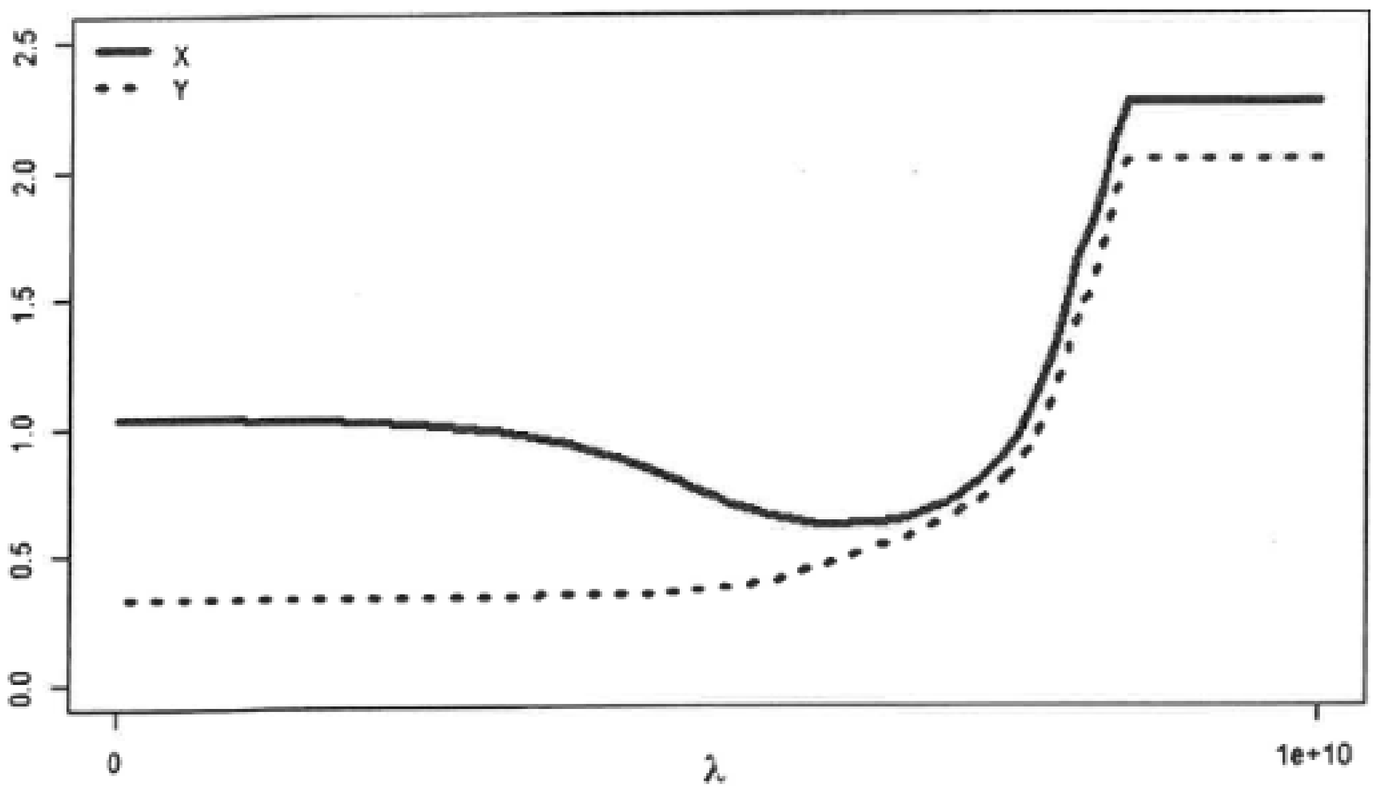
36. A modeler creates a local regression model.
After reviewing the results, the fitted line appears too wiggly, over-responding to trends in nearby data points.
The modeler would like to adjust the model to produce more intuitive results.
Determine which one of the following adjustments the modeler should make.
A. Add a linear constraint in the regions before and after the first knot
B. Increase the number of orders in the regression equation
C. Increase the number of knots in the model
D. Reduce the number of knots in the model
E. Increase the span, s, of the model

36. E. Increasing the span, s, of the model, will include a larger percentage of the observations in each local regression, reducing the effect of a few points.
<u>Comment</u>: There are no knots in a local regression. Splines have knots.
Local regression fits a series of <u>straight</u> lines via weighted regression; in any case, increasing the number of orders in the regression equation would introduce more wiggling.

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

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

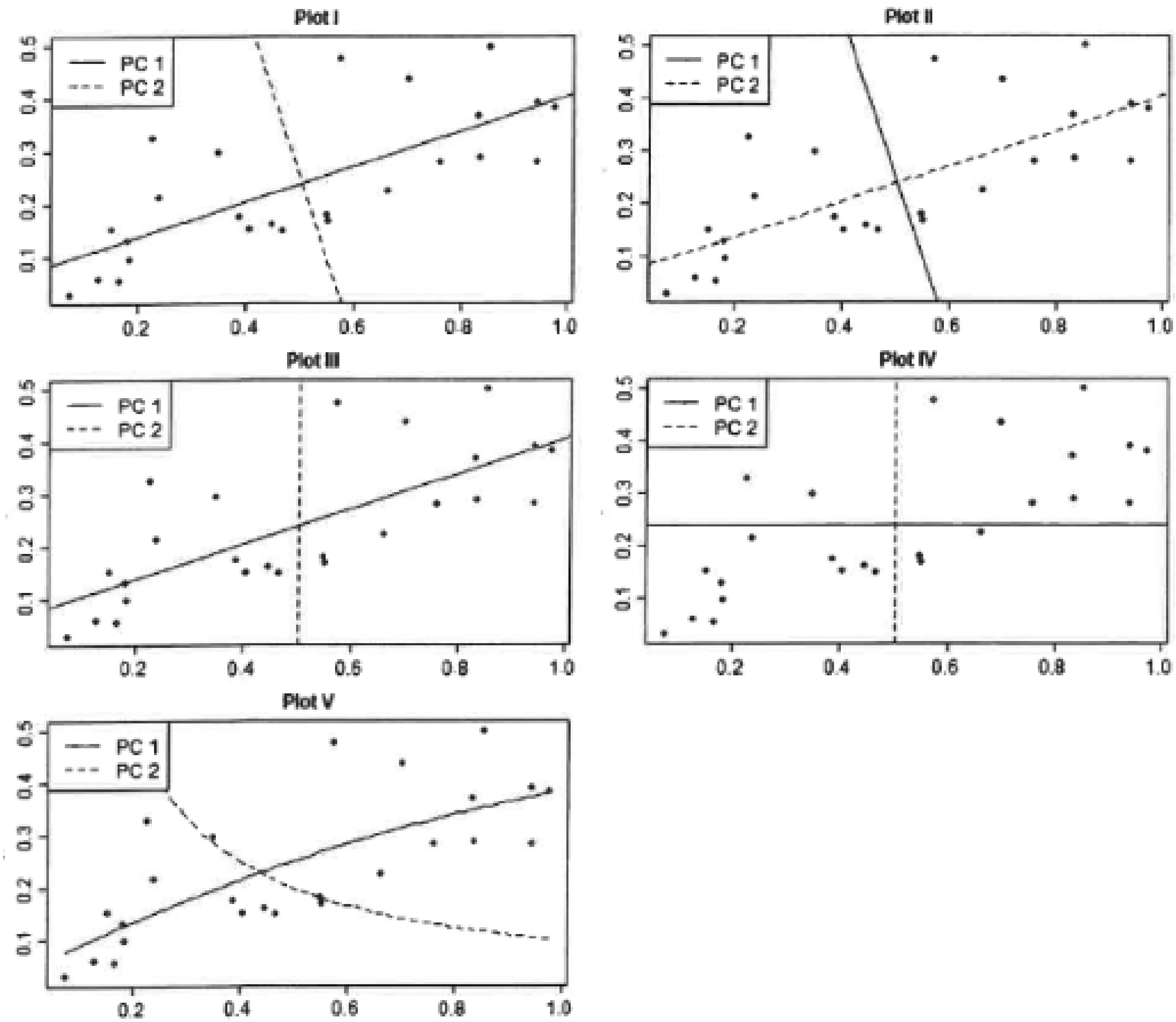
From this model you have produced the following plot of various statistics as a function of tuning parameters, λ :



Determine which of the following statistics X and Y represent.
A. X = Squared Bias, Y = Training MSE
B. X = Test MSE, Y = Training MSE
C. X = Test MSE, Y = Variance
D. X = Training MSE, Y = Variance
E. X = Variance, Y = Test MSE

37. B. The Test MSE has a U-shaped curve. $\Rightarrow$ X = Test MSE.
The function to be minimized has the form of the lasso.
As λ increases, the number of non-zero slopes decreases, until eventually we have the null model with no predictors. As we have fewer predictors, we are less able to fit the data, and the training MSE increases. $\Rightarrow$ Y = Training MSE.
Comment: See Figure 6.8 in An Introduction to Statistical Learning.
As λ increases, the effect of the penalty increases, and the flexibility of the fitted model decreases. For lambda = 0 we get the least squares model, which has no bias but a large variance. As lambda increases, the squared bias increases while the variance decreases. Thus Y could be the squared bias, but that is not a choice.

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



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

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

38. A. The principal components are orthogonal (perpendicular to each other), eliminating III and V.
Principal component 1 should be in the direction of the most variation between the data points; that is true for **Plot I**.
(Principal component 2 is then the direction that picks up the remaining variation.)
Comment: See Figure 6.14 in An Introduction to Statistical Learning.
With only two variables, there are only two principal components.
The two lines should cross at $(\bar{X}, \bar{Y})$.

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

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

Below are statistics for all candidate models:

Model	# of Non Intercept Parameters	Parameters	R^2	Log-likelihood
1	0	I	0	0.05
2	1	I, 1	0.56	1.3
3	1	I, 2	0.57	1.4
4	1	I, 3	0.55	1.2
5	1	I, 4	0.52	1.15
6	1	I, 5	0.51	1.1
7	2	I, 1, 2	0.61	2.5
8	2	I, 1, 3	0.64	2.75
9	2	I, 1, 4	0.63	2.6
10	2	I, 1, 5	0.69	3
11	2	I, 2, 3	0.61	2.5
12	2	I, 2, 4	0.62	2.55
13	2	I, 2, 5	0.68	2.9
14	2	I, 3, 4	0.66	2.8
15	2	I, 3, 5	0.64	2.75
16	2	I, 4, 5	0.6	2.45
17	3	I, 1, 2, 3	0.73	3.35
18	3	I, 1, 2, 4	0.71	3.25
19	3	I, 1, 2, 5	0.72	3.3
20	3	I, 1, 3, 4	0.75	3.5
21	3	I, 1, 3, 5	0.76	3.6
22	3	I, 1, 4, 5	0.79	3.9
23	3	I, 2, 3, 4	0.78	3.7
24	3	I, 2, 3, 5	0.74	3.4
25	3	I, 2, 4, 5	0.75	3.45
26	3	I, 3, 4, 5	0.73	3.35
27	4	I, 1, 2, 3, 4	0.88	4.2
28	4	I, 1, 2, 3, 5	0.8	3.95
29	4	I, 1, 2, 4, 5	0.87	4.1
30	4	I, 1, 3, 4, 5	0.83	4
31	4	I, 2, 3, 4, 5	0.85	4.05
32	5	I, 1, 2, 3, 4, 5	0.9	4.25

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

Calculate the absolute value of the difference between AIC_A and AIC_B.

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

39. B. AIC = (-2) (loglikelihood) + (number of parameters)(2).
In each case, I will include the intercept in the number of parameters; this does not affect the comparisons as long as one is consistent.
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	Log-Likelihood	AIC
1	1	0.05	1.9
3	2	1.4	1.2
10	3	3	**0**
22	4	3.9	0.2
27	5	4.2	1.6
32	6	4.25	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	Log-Likelihood	AIC
1	I	1	0.05	1.9
3	I, 2	2	1.4	1.2
13	I, 2, 5	3	2.9	**0.2**
25	I, 2, 4, 5	4	3.45	1.1
29	I, 1, 2, 4, 5	5	4.1	1.8
32	I, 1, 2, 3, 4, 5	6	4.25	3.5

Model 13 has the best (smallest) AIC at 0.2. $AIC_B = 0.2$.
$|AIC_A - AIC_B| = |0 - 0.2|$ = **0.2**.
<u>Comment</u>: When one has ft all of the possible models as here, Best Subset Selection does at least as well as Forward Stepwise Regression, since the latter only considers some of the possible models. The path chosen by Forward Stepwise Regression might miss the very best model, as was the case here. The advantage of Forward Stepwise Regression is that we would fit fewer models; this makes a huge difference when there are many more than 5 predictors.
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	Log-Likelihood	AIC
32	I, 1, 2, 3, 4, 5	6	4.25	3.5
27	I, 1, 2, 3, 4	5	4.2	1.6
23	I, 2, 3, 4	4	3.7	0.6
14	I, 3, 4	3	2.8	**0.4**
4	I, 3	2	1.2	1.6
1	I	1	0.05	1.9

40. A model is created with the following form:

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

where: $(x - \xi)_+^3 = (x - \xi)^3$ if $x > \xi$, and 0 otherwise

You are given the following statements:

I. f(x) is continuous at ξ

II. f'(x) is continuous at ξ

III. f''(x) is continuous at ξ

Determine which of the above statements are true.

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

40. D. The cubic polynomial portion is continuous, and has continuous first and second derivatives. As x approaches ξ from the right, $(x - \xi)_+^3 = (x - \xi)^3$ approaches zero.

Thus f(x) is continuous at ξ.

The derivative of $(x - \xi)^3$ is $3(x - \xi)^2$, which goes to zero as x approaches ξ from the right.

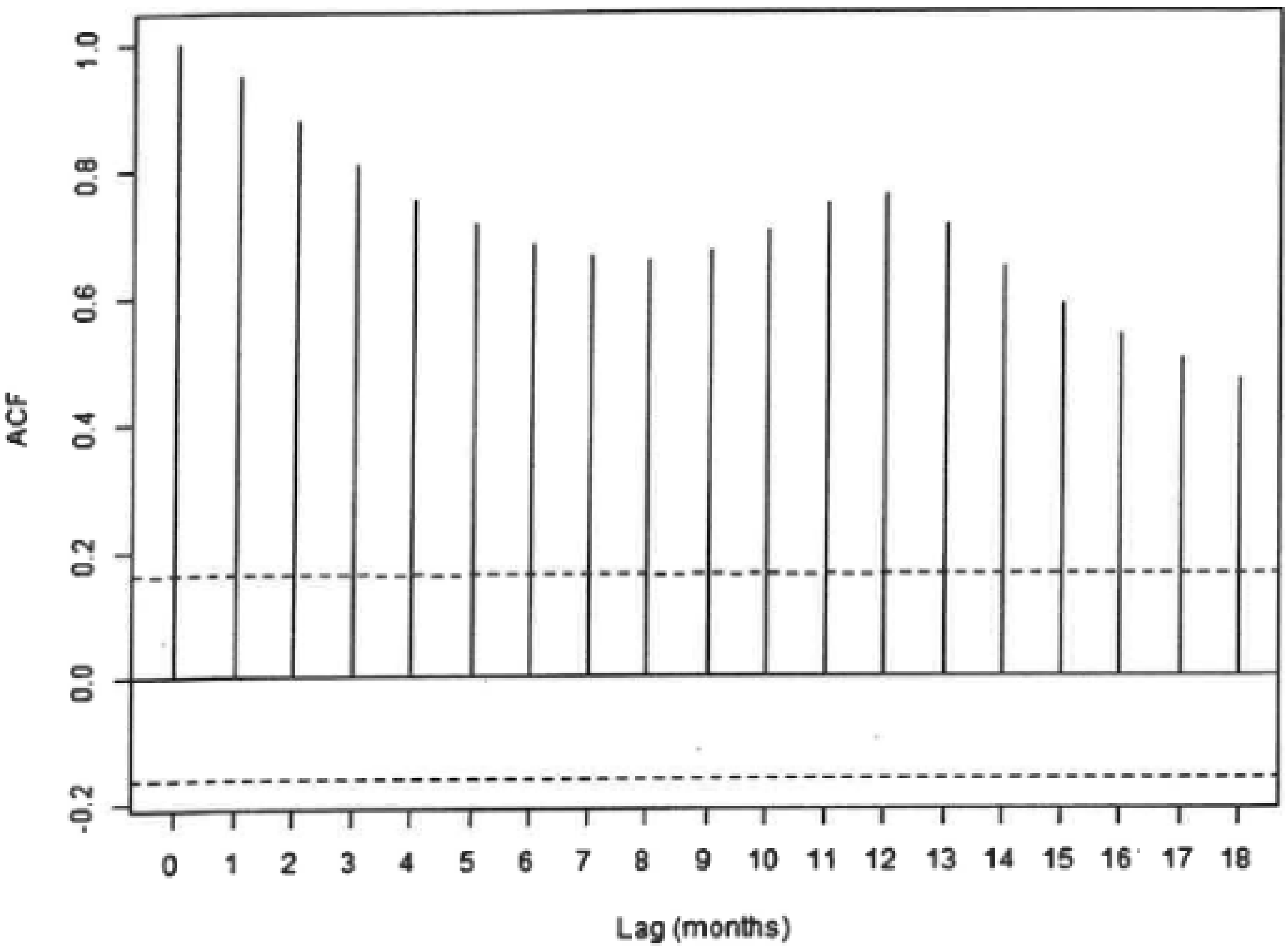
Thus f'(x) is continuous at ξ.

The second derivative of $(x - \xi)^3$ is $6(x - \xi)$, which goes to zero as x approaches ξ from the right.

Thus f''(x) is continuous at ξ.

<u>Comment</u>: This is the basis form for a cubic spline with one knot at ξ.

41.
An actuary produces the following correlogram for vehicle accident severities over a 10-year period:



You are also given the following three statements about this time series:
I. There is a positive trend in accident severity
II. There is a negative trend in accident severity
III. The accident severity data shows a seasonal pattern
Determine which of the above statements can be conclude from the above graph.
A. I only
B. II only
C. III only
D. I, II and III
E. The answer is not given by (A), (B), (C) or (D)

41. C. The correlogram seems to be a cosine function imposed on an approximately linear decline.
This indicates a cycle plus a trend.
The cycle appears to be a 12-month cycle, in other words a seasonal pattern.
While the correlogram indicates that there is a trend, there is no way to conclude whether any such trend is positive or negative.
Comment: See Figure 2.6 in Introductory Time Series.
While typically accident severities show an upward trend over a 10-year period due to inflation, one can not infer that from the given information.
The final sentence of the question should have read “can be conclude**d** from”.

42. A time series is modeled using the function below:

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

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

Calculate the forecast value of x_7.

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

42. D. $\nabla x_3 = -1.3. \Rightarrow x_3 - x_2 = -1.3. \Rightarrow x_2 = x_3 + 1.3 = 45.3 + 1.3 = 46.6.$

$\nabla x_2 = 26.2. \Rightarrow x_2 - x_1 = 26.2. \Rightarrow x_1 = x_2 - 26.2 = 46.6 - 26.2 = 20.4.$

$x_1 = \alpha_0 + \alpha_1 + z_1. \Rightarrow 20.4 = \alpha_0 + \alpha_1 - 9.6. \Rightarrow 30 = \alpha_0 + \alpha_1.$

$x_3 = \alpha_0 + 3\alpha_1 + z_3. \Rightarrow 45.3 = \alpha_0 + 3\alpha_1 - 4.7. \Rightarrow 50 = \alpha_0 + 3\alpha_1.$

Subtracting the two equations: $20 = 2\alpha_1. \Rightarrow \alpha_1 = 10. \Rightarrow \alpha_0 = 20.$

$\hat{x}_7 = \alpha_0 + 7\alpha_1 = 20 + (7)(10) = \mathbf{90}.$

Comment: The difference operator is defined as: $\nabla x_t = x_t - x_{t-1}$.

Backwards question; we need to solve for the slope and intercept of the linear model.
z_t is a white noise series. $\Rightarrow z_7$ is independent of the previous observed values of z_t, and $E[z_7] = 0$.

43. You are given the following time series model:

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

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

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

A. Non-stationary, invertible

B. Non-stationary, not invertible

C. Stationary, invertible

D. Stationary, not invertible

E. The correct answer cannot be determined from the information given

43. A. $\theta(B) = 1 - \frac{2}{3}B - \frac{1}{3}B^2$.

Find the roots: $0 = 1 - \frac{2}{3}B - \frac{1}{3}B^2$. $\Rightarrow$ $B = \dfrac{2/3 \pm \sqrt{(-2/3)2 - (4)(-1/3)(1)}}{(2)(-1/3)}$ = 1 or -3.

Since not all of the roots have an absolute value greater than one, the time series is <u>not</u> stationary.

We can write the time series as: $\theta(B)\, x_t = \phi(B) w_t$, with $\phi(B) = 1$.

We can rewrite this as: $\phi^{-1}(B)\ \theta(B)\ x_t = w_t$, with $\phi^{-1}(B) = 1$.

Thus this time series is invertible.

Alternately, set $\phi(B) = 0$. $\Rightarrow 1 = 0$. There are no roots. Thus there are no roots of absolute value of one or less. $\Rightarrow$ This time series is invertible.

<u>Comment</u>: The concept of invertibility is usually only applied to MA and ARMA models, while this is an AR(2) model. However, an AR(p) model is also an ARMA(p, 0) model.

In this sense. all AR(p) series are invertible.

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

I. When there is positive serial correlation in a time series, the standard errors of the estimated regression parameters are likely to be over-estimated.

II. GLS is an improvement over ordinary least squares regression for serially correlated time series because GLS is based on maximizing the likelihood given the white noise in the data.

III. GLS can be used to provide better estimates of standard errors of the regression parameters to account for autocorrelation in the residual series.

Determine which of the above statements are true.

A. I only

B. II only

C. III only

D. I, II and III

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

44. C. Statement I is backwards. When there is positive serial correlation in a time series, when fitting via least squares, the standard errors of the estimated regression parameters are likely to be underestimated, and vice-versa for negative serial correlation.

Statement II is not true. The correct statement does not involve the white noise, which is random, but rather than the series of residuals which are usually correlated in the case of time series.

Statement II is True.

Comment: See Section 5.4 of Cowpertwait and Metcalfe.

"Generalized Least Squares (GLS) can be used to provide better estimates of the standard error of the regression parameters to account for the autocorrelation of the residual series. The procedure is essentially based on maximizing the likelihood given the autocorrelation in the data."

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

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

Calculate the sample lag 4 autocorrelation.

A. Less than 0.0

B. At least 0.0, but less than 0.3

C. At least 0.3, but less than 0.6

D. At least 0.6, but less than 0.9

E. At least 0.9

45. A. The mean is 24.

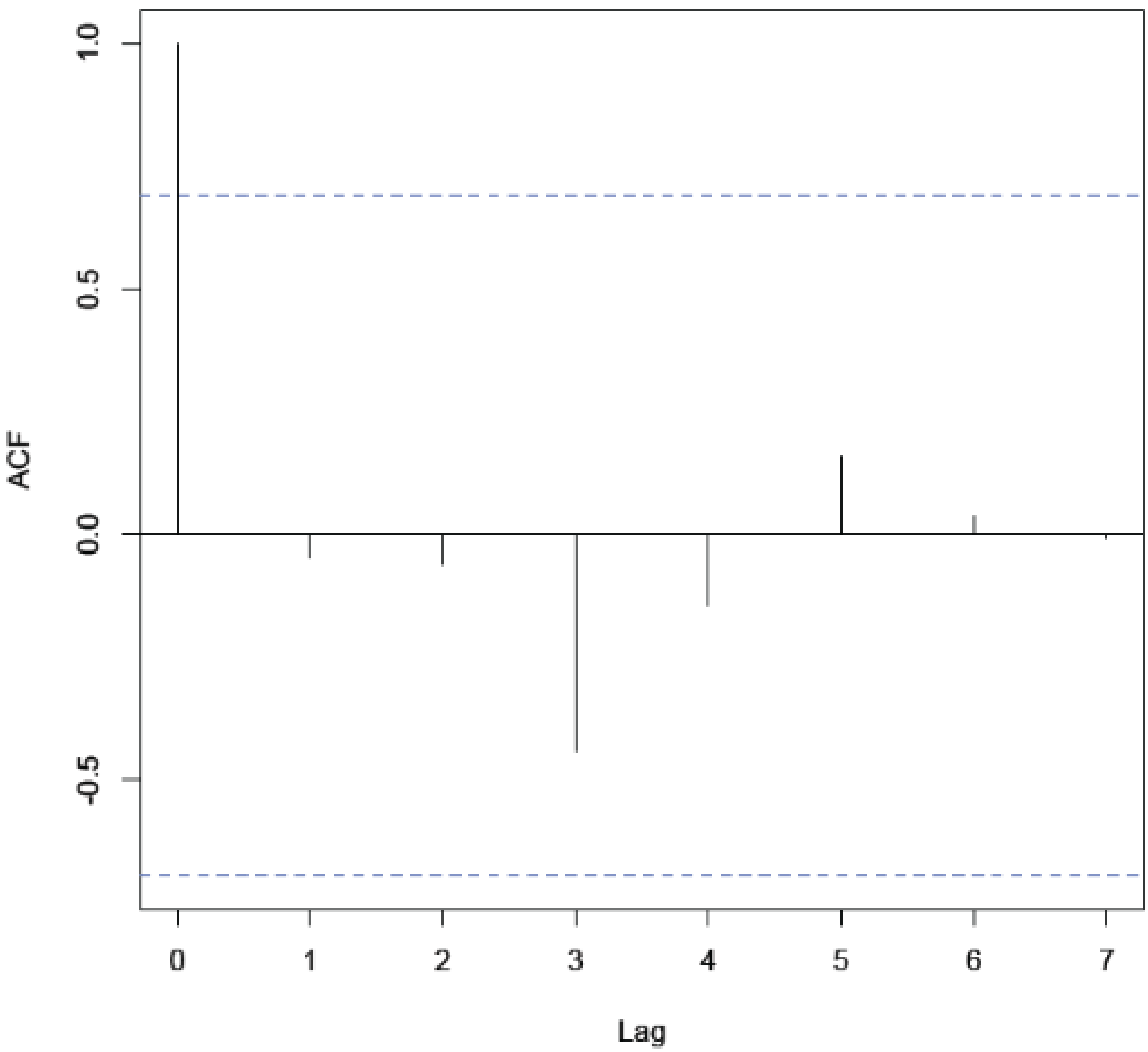
$$c_0 = \frac{(25-24)^2 + (19-24)^2 + (10-24)^2 + (32-24)^2 + (26-24)^2 + (38-24)^2 + (22-24)^2 + (20-24)^2}{8}$$

$= 63.25.$

$$c_4 = \frac{(25-24)(26-24) + (19-24)(38-24) + (10-24)(22-24) + (32-24)(20-24)}{8} = -9.$$

$r_4 = c_4/c_0 = -9/63.25 =$ **-0.14**.

Comment: The correlogram:



END OF EXAMINATION