CA – Exam 4

Exam Results

Nov 7th, 2024 at 12:28 AM

Back to Practice

Score 35/45	Difficulty	Earned Level	Total Time
78%	4.7	5.2 +1.0	4:52:12 of 4:00:00

Section Review

1 Probability Models	12/16	75%	2:02:21	36%
2 Statistics	6/8	75%	49:18	18%
3 Extended Linear Models	17/21	81%	2:00:33	47%

1/1 | 63% | 1.1 | 4.8 | 6:16 | 3:37

1
2
3
4
5
6
7
8
9
10
11
12

You are given the following information about a sample, X_1, ..., X_n:

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

Calculate the probability that $Y > \theta$.

4% A Less than 0.2

63% At least 0.2, but less than 0.4

21% C At least 0.4, but less than 0.6

10% D At least 0.6, but less than 0.8

3% E At least 0.8

1 1/1 66% 1.4 4.8 4:34 5:06

1
2
3
4
5
6
7
8
9
10
11
12

You are given the following information:

- Coins are tossed into a fountain according to a Poisson process at a rate of one every two minutes.
- The coin denominations are independently distributed as follows:

Coin Denomination	Probability
Penny	0.50
Nickel	0.20
Dime	0.20
Quarter	0.10

Calculate the probability that two Quarters will be tossed into the fountain before four non-Quarter coins.

8% A Less than 0.06

6% B At least 0.06, but less than 0.07

10% C At least 0.07, but less than 0.08

66% D At least 0.08, but less than 0.09

9% E At least 0.09

1/1 67% 1.1 3.3 9:44 4:08

1
2
3
4
5
6
7
8
9
10
11
12

You are given the following information about a watch with 6 different parts:

- There are 3 red wires with expected lifetimes of 50, 75, and 100.
- There are 3 yellow wires with expected lifetimes of 25, 50, and 75.
- The lifetimes of all wires are independent and exponentially distributed.

Calculate the probability that a red wire will break down before a yellow wire.

4% A Less than 0.20

11% B At least 0.20, but less than 0.25

11% C At least 0.25, but less than 0.30

8% D At least 0.30, but less than 0.35

67% At least 0.35

1/1 66% 1.4 4.1 6:19 5:25

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given the following information:

- Taxicabs leave a hotel with a group of passengers at a Poisson rate of 10 per hour.
- The number of people in each group taking a cab is independent and has the following probability:

Number of People	Probability
1	0.60
2	0.25
3	0.10
4	0.05

Calculate the probability that at least 800 people leave the hotel in a cab during a 48-hour period using the normal approximation.

12% A Less than 0.200

13% B At least 0.200, but less than 0.210

66% At least 0.210, but less than 0.220

3% D At least 0.220, but less than 0.230

6% E At least 0.230

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given the following information about emergency room visits at a local hospital:

- The amount of time between emergency room visits for broken bones is exponentially distributed with a mean of two hours.
- The amount of time between emergency room visits for the flu is exponentially distributed with a mean of five hours.
- Emergency room visits for broken bones and emergency room visits for the flu are independent.

Calculate the probability that two patients will come into the hospital with broken bones before three patients come into the hospital with the flu.

14% A Less than 0.875

7% B At least 0.875, but less than 0.900

9% C At least 0.900, but less than 0.925

64% ✓ At least 0.925, but less than 0.950

6% E At least 0.950

0/1 62% 1.5 5.9 14:26 2:56

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given the minimal cut sets of a system:

$$\{1\} \quad \{2\} \quad \{3,4\} \quad \{4,5\}$$

Determine the number of minimal path sets this system has.

Incorrect Answer

62% Fewer than 3

18% B 3

15% 4

3% D 5

3% E Greater than 5

1/1 66% 1.5 4.8 5:52 3:30

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given the following information:

- System X is a series system with two components.
- System Y is a parallel system with two components.
- All components are independent and function for an amount of time, uniformly distributed over (0,1).

Calculate the absolute value of the difference in expected system lifetimes between system X and system Y.

3% A Less than 0.10

7% B At least 0.10, but less than 0.20

9% C At least 0.20, but less than 0.30

66% At least 0.30, but less than 0.40

15% E At least 0.40

0/1 58% 1.5 6.2 5:55 3:55

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given the following information on a random graph:

- The graph has six nodes.
- The probability that there is an arc between two nodes is 0.6.
- Every arc is independent.

Calculate the probability that the graph is connected.

Incorrect Answer

6% A Less than 0.900

10% At least 0.900, but less than 0.925

58% At least 0.925, but less than 0.950

17% D At least 0.950, but less than 0.975

8% E At least 0.975

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given the following Markov chain transition matrix:

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 & 0.0 & 0.0 \\ 0.4 & 0.4 & 0.2 & 0.0 & 0.0 \\ 0.5 & 0.5 & 0.0 & 0.0 & 0.0 \\ 0.4 & 0.2 & 0.1 & 0.3 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

Determine the number of transient states in this Markov chain.

1 (round to the nearest whole number)

Correct Answer: 1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given the following probability transition matrix for a Markov chain with three states labeled 0, 1, and 2:

$$P = \begin{bmatrix} 0.5 & c & 0.5-c \\ 0.3 & 0.6 & 0.1 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}$$

The long-run probability that the chain is in state 0 is 0.3415.

Calculate the long-run probability that the chain is in state 2.

Incorrect Answer

64% Less than 0.28

5% B At least 0.28, but less than 0.30

11% C At least 0.30, but less than 0.32

7% D At least 0.32, but less than 0.34

13% At least 0.34

1/1 74% 1.6 2.6 2:39 3:39

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given the following transition matrix for a Markov chain with two states 0,1:

$$A = \begin{bmatrix} 0.25 & 0.75 \\ 0.01 & 0.99 \end{bmatrix}$$

At time $t = 0$, the Markov chain is in state 0.

Calculate the expected number of steps needed to return to state 0.

7% A Fewer than 10

7% B At least 10, but fewer than 30

7% C At least 30, but fewer than 50

5% D At least 50, but fewer than 70

74% At least 70

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

A life annuity has the following values:

- $\ddot{a}_{20} = 13.9294$
- $\ddot{a}_{21} = 13.8529$
- The interest rate used is 0.05.
- The force of mortality is constant between the ages of 20 and 30.

From a sample of 500 independent twenty-year-olds, you wish to calculate a 95% confidence interval for the number of people who will die within five years implied by the annuity values above, using a normal approximation.

Calculate the upper bound of this confidence interval.

17% A Less than 60.0

58% B At least 60.0, but less than 65.0

14% C At least 65.0, but less than 70.0

6% D At least 70.0, but less than 75.0

5% E At least 75.0

1/1 70% 1.7 3.5 8:46 4:43

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

A company plans to offer a warranty on their product, which pays 1,000 at the end of the year of failure if their product fails within a certain number of years, N. You are given the following information:

- The annual interest rate is 5%.
- The rates of failure, q, for each year are given by the following table:

Age x	q_x
0	0.01
1	0.03
2	0.10
3	0.15
4	0.30

- The length of the warranty, N, is set to be the largest number of years such that the actuarial present value of the warranty is less than 200.

Calculate N.

1% A 1

17% B 2

70% 3

7% D 4

5% E 5

1/1 65% 1.7 4.8 1:30 5:10

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

An insurance company is using the Illustrative Life Table to price a block of life insurance policies covering 5,000 people aged 50.

Calculate the lower bound of a 90% confidence interval for the number of deaths in this block during the next 15 years, using a normal approximation.

- 4% A Less than 745
- 65% At least 745, but less than 755
- 16% C At least 755, but less than 765
- 7% D At least 765, but less than 775
- 9% E At least 775

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given:

- X is a beta random variable with $a = 2, b = 2$ and $\theta = 2$.
- X is simulated using the acceptance-rejection method with the following random variable with density function:

$$g(y) = \frac{1}{2}, \text{ where } 0 < y < 2$$

- The following values from the random variable above are given: $0.1, 0.6, 0.4, 1.3, 1.7$.
- The following values from the uniform distribution [0,1] are given: $0.71, 0.22, 0.59, 0.08, 0.55$.

Calculate the first simulated value of X.

9% A 0.1

77% 0.6

8% C 0.4

5% D 1.3

1% E 1.7

0/1 | 78% | 1.8 | 1.9 | 17:43 | 3:47

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Uniform random numbers are simulated using a mixed congruential generator

$$x_{n+1} = (ax_n + c) \bmod m$$

with the following parameters:

$$\begin{aligned} a &= 77 && \text{(the multiplicative factor)} \\ c &= 643 && \text{(the additive term)} \\ m &= 4{,}096 && \text{(the modulus)} \end{aligned}$$

The initial seed, x_0, is not known but it is given that the second generated pseudorandom value, x_2, is 3,623.

Determine the sixth generated uniform random number.

Incorrect Answer

4% A Less than 0.2

7% B At least 0.2, but less than 0.4

9% C At least 0.4, but less than 0.6

78% D At least 0.6, but less than 0.8

2% E At least 0.8

1/1 | 69% | 2.1 | 3.7 | 1:19 | 3:36

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Buses arrive at a bus stop according to a Poisson process with parameter λ. Starting from time $t = 0$, you observe bus inter arrival times of $t = 7, 3, 5, 3, 2$.

Calculate the maximum likelihood estimate of λ.

4% A Less than 0.1

3% B At least 0.1, but less than 0.2

69% At least 0.2, but less than 0.3

15% D At least 0.3, but less than 0.4

9% E At least 0.4

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given the following information:

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

Calculate c.

65% Less than 1.5

7% B At least 1.5, but less than 2.5

9% C At least 2.5, but less than 3.5

16% D At least 3.5, but less than 4.5

3% E At least 4.5

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

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

The true value of θ is 10.

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

Calculate N.

12% A Less than 20

16% B At least 20, but less than 40

15% C At least 40, but less than 60

4% D At least 60, but less than 80

54% At least 80

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given:

- A random sample, $X_1, X_2, \ldots, X_{10}$ from the Bernoulli distribution with $q = \Pr(X = 1)$.
- $H_0 : q = 0.5$ and $H_1 : q > 0.5$.
- The critical region, $C = \left\{ \sum_{k=1}^{10} x_k > 6 \right\}$.

Calculate the probability of Type I error.

3% A Less than 0.1

67% At least 0.1, but less than 0.2

17% C At least 0.2, but less than 0.3

10% D At least 0.3, but less than 0.4

3% E At least 0.4

0/1 66% 2.4 5.2 7:55 5:21

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Suppose that the number of claims filed by a given customer, X, is such that $X = 0$ with probability p and X has a Poisson distribution with mean 3 with probability $1 - p$.

You want to perform the hypothesis test

$$H_0 : p = 0.1 \quad \text{vs.} \quad H_A : p = 0.2$$

You take a single observation of X and reject the null hypothesis when $X = 0$.

Consider the following statements regarding this hypothesis test:

I. The probability of a Type I error is less than 0.2.
II. The probability of a Type II error is less than 0.2.
III. This is the most powerful test of these hypotheses at its significance level.

Determine which of the above statements about this hypothesis test is/are correct.

Incorrect Answer

7% None is correct

9% B I and II only

66% I and III only

7% D II and III only

9% E The answer is not given by (A), (B), (C), or (D).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

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.

10% A Less than 17

56% ✓ 17

17% C 18

5% D 19

13% E At least 20

1/1 67% 2.6 4.7 10:52 6:23

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given:

- Ten independent observations which follow the normal distribution with unknown mean μ and variance $\sigma^2 = 20$.
- Based on your observations, the $(1-\alpha)$ confidence interval for μ is $(-6.84, -1.30)$.
- $H_0 : \mu = -5$ and $H_1 : \mu \neq -5$

Calculate the critical region at significance level α.

9% A $(-\infty, -6.84) \cup (6.84, \infty)$

3% B $(-6.84, 6.84)$

15% C $(-7.77, -2.23)$

67% ✔ $(-\infty, -7.77) \cup (-2.23, \infty)$

7% E $(-\infty, -10.54) \cup (0.54, \infty)$

0/1 60% 2.7 5.7 9:33 4:24

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

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

The density for the k^{th} order statistic of a sample is given as:

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

Calculate the expected bias of this estimate.

Incorrect Answer

12% A $-\frac{4\theta}{5}$

8% B $-\frac{4\theta}{7}$

60% C $-\frac{\theta}{10}$

2% D $-\frac{\theta}{30}$

19% E The answer is not given by (A), (B), (C), or (D).

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

You fit a least squares regression to five pairs of observations (x_i, y_i) using the following model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

You determine:

- $\sum_{i=1}^{5} x_i = 10$
- $\sum_{i=1}^{5} x_i^2 = 30$
- $\sum_{i=1}^{5} \left(y_i - \hat{y}_i\right)^2 = 15$

Calculate the estimated variance of the estimator $\hat{\beta}_1$.

12% A Less than 0.4

66% ✔ At least 0.4, but less than 0.8

12% C At least 0.8, but less than 1.8

6% D At least 1.8, but less than 2.8

4% E At least 2.8

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

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.

4% A Less than 1

74% ✔ At least 1, but less than 3

12% C At least 3, but less than 5

7% D At least 5

3% E The answer cannot be computed from the information given.

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

You are given the following linear regression model which is fitted to 11 observations:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

The coefficient of determination is $R^2 = 0.25$.

Calculate the F-statistic used to test for a linear relationship.

8% A Less than 1.5

10% B At least 1.5, but less than 2.5

78% C At least 2.5, but less than 3.5

3% D At least 3.5, but less than 4.5

2% E At least 4.5

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

You are given the following information regarding a GLM that was used to predict claim severity of collision coverage in auto insurance:

- The available predictors are:
  - Number of Drivers ("1", "2", "3+")
  - Territory ("A", "B", "C", "D", "E")
- 15 observations are used in the model.
- The partial ANOVA table is given below:

Response variable	Losses
Response distribution	Normal
Link	Identity

Source of variation	Degrees of freedom	Sum of squares
Number of Drivers		28,538.12
Territory		109,058.85
Residuals		133,520.43

- You want to evaluate whether or not the predictor Territory is significant in predicting Losses using an F-test.

Calculate the p-value of this test.

3% A. Less than 2%

4% B. At least 2%, but less than 5%

6% C. At least 5%, but less than 10%

12% D. At least 10%, but less than 20%

75% E. At least 20%

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

You are given the following table of eight individual adults' heights in inches, y_i, categorized by their season of birth, and selected summary statistics:

Observation	Height	Season
1	70	Summer/Fall
2	67	Summer/Fall
3	64	Summer/Fall
4	66	Summer/Fall
5	60	Winter/Spring
6	62	Winter/Spring
7	68	Winter/Spring
8	63	Winter/Spring

- $\bar{y} = 65$
- Observation is indexed by i.
- $\sum_{j=1}^{8} y_i^2 = 33{,}878$
- $\left(\sum_{j=1}^{8} y_i\right)^2 = 270{,}400$
- $\sum_{i=1}^{4} (y_i - \bar{y}_1)^2 = 18.75$ where $\bar{y}_1$ is the mean of heights for individuals born in summer or fall
- $\sum_{j=5}^{8} (y_i - \bar{y}_2)^2 = 34.75$ where $\bar{y}_2$ is the mean of heights for individuals born in winter or spring

Calculate the mean square for error of the observations.

9% A Less than 7.5

11% B At least 7.5, but less than 8.5

60% C At least 8.5, but less than 9.5

8% D At least 9.5, but less than 10.5

12% E At least 10.5

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

You are fitting a linear regression model of the form:

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

and are given the following values used in this model:

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

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

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

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

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

7% A 0

65% ✓ 1

17% C 2

5% D 3

7% E 4

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

A linear model is fitted to a data set with the following observations:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

i	y_i	$x_{i,1}$	$x_{i,2}$
1	0.8226	0.5246	0.4204
2	0.7210	0.9704	0.6895
3	0.9617	0.8083	0.2000
4	0.2221	0.1827	0.3401
5	0.1013	0.3135	0.5923

The results are as given below:

- $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \begin{pmatrix} 0.4171 \\ 1.0068 \\ -0.9256 \end{pmatrix}$
- $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \begin{pmatrix} 0.2069 & 0.1337 & 0.2249 & 0.2423 & 0.1922 \\ 0.1337 & 0.8564 & 0.0672 & -0.2511 & 0.1938 \\ 0.2249 & 0.0672 & 0.8449 & 0.1139 & -0.2509 \\ 0.2423 & -0.2511 & 0.1139 & 0.5609 & 0.3340 \\ 0.1922 & 0.1938 & -0.2509 & 0.3340 & 0.5309 \end{pmatrix}$

Calculate the leave-one-out cross validation (LOOCV) estimate for the test MSE.

13% A Less than 0.10

61% B At least 0.10, but less than 0.20

14% C At least 0.20, but less than 0.30

7% D At least 0.30, but less than 0.40

5% E At least 0.40

0/1 65% 3.6 5.3 4:19 2:34

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Consider the following statements concerning a data set with n observations and p predictors:

I. Performing best subset selection requires fitting all $\binom{p}{2}$ possible models.

II. If $n = 100$ and $p = 4$, performing forward stepwise selection requires fitting 11 distinct models.

III. Backward stepwise selection cannot be performed on the data set if $n < p$.

Determine which of the statements is/are true.

Incorrect Answer

6% A I and II only

9% B I and III only

65% II and III only

7% D I, II, and III

14% The correct answer is not given by (A), (B), (C), or (D).

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38

Consider the following statements:

I. The proportion of variance explained by an additional principal component increases as more principal components are added.
II. The cumulative proportion of variance explained increases as more principal components are added.
III. Using all possible principal components provides the best understanding of the data.

Determine which of the statements are correct.

5% A I only

75% II only

6% C III only

8% D I, II, and III

6% E The correct answer is not given as (A), (B), (C), or (D).

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

Bryce is a pricing actuary in Auto City Insurance Company. He performs a principal component analysis on a data set with the following variables:

- X_1 : Number of major and minor violations in the last year
- X_2 : Credit score

All these variables are centered, so they have means of 0. The results of the principal component analysis are as follows:

- The first principal component loading for X_1 is 0.8.
- Both the first and second principal component loading for X_2 are negative.

Calculate the second principal component score for an observation with $X_1 = -1.0$, and $X_2 = 1.8$.

19% A Less than -1.0

65% B At least -1.0, but less than -0.5

9% C At least -0.5, but less than 0.0

4% D At least 0.0, but less than 0.5

3% E At least 0.5

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

For a set of data with 40 observations, 2 predictors (X_1 and X_2), and one response (Y), the residual sum of squares has been calculated for several different estimates of a linear model with no intercept. Only integer values from 1 to 5 were considered for estimates of β_1 and β_2.

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

		$\hat{\beta}_2$				
		1	2	3	4	5
	1	2,855.0	870.3	464.4	357.2	548.6
	2	1,059.1	488.4	216.3	242.8	567.9
$\hat{\beta}_1$	3	657.0	220.0	81.6	241.9	700.8
	4	368.4	65.1	60.5	354.5	947.1
	5	193.2	23.7	152.8	580.6	1,307.0

Let:

- $\hat{\beta}_1^L$ = Estimate of β_1 using a lasso with budget parameter $s = 5$
- $\hat{\beta}_2^L$ = Estimate of β_2 using a lasso with budget parameter $s = 5$

Calculate the ratio $\hat{\beta}_1^L / \hat{\beta}_2^L$.

Incorrect Answer

2% A Less than 0.5

63% At least 0.5, but less than 1.0

13% C At least 1.0, but less than 1.5

7% D At least 1.5, but less than 2.0

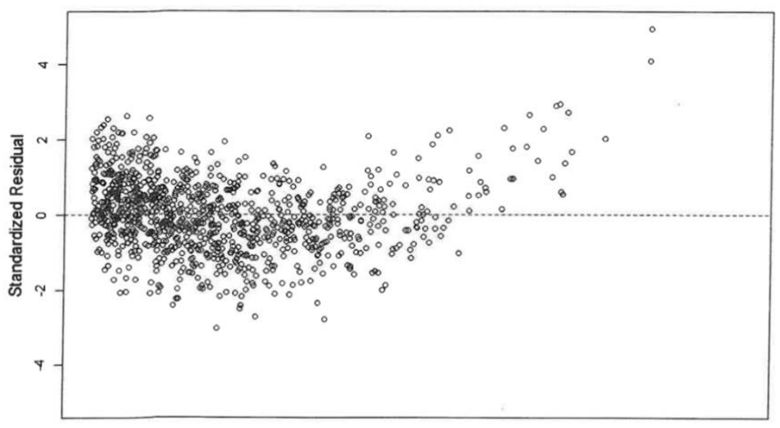
16% At least 2.0

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

A GLM has been fit to a data set with 1,000 observations with the following model form:

$$\mathrm{E}(y_i) = g^{-1}(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})$$

The standardized residuals are plotted against the values of the variable represented by X_3 in the model. The plot is shown below.



Determine the best alternate model parameterization based on the residual plot above.

72% $\mathrm{E}(y_i) = g^{-1}(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{3i}^2)$

2% B $\mathrm{E}(y_i) = g(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})$

4% C $\mathrm{E}(y_i) = g^{-1}(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_4 X_{2i}^2 + \beta_3 X_{3i})$

10% D $\mathrm{E}(y_i^2) = g^{-1}(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})$

13% E $\mathrm{E}(\ln y_i) = g^{-1}(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})$

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

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.

18% A Less than 0.15

71% At least 0.15, but less than 0.30

7% C At least 0.30, but less than 0.45

2% D At least 0.45, but less than 0.60

2% E At least 0.60

1/1 | 60% | 3.8 | 6.2 | 11:57 | 4:23

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

You are given that X follows a negative binomial distribution with parameters r and β. One observation x from X is observed.

Determine the score function that corresponds to β.

3% A $\dfrac{x\beta}{1+\beta}$

16% B $\dfrac{x}{\beta} - \dfrac{r-x}{1+\beta}$

10% C $\dfrac{x}{\beta(1+\beta)} + \dfrac{r}{1+\beta}$

12% D $\dfrac{x+r\beta}{\beta(1+\beta)}$

60% $\dfrac{x-r\beta}{\beta(1+\beta)}$

1/1 | 71% | 3.8 | 3.7 | 4:21 | 3:28

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

You are given the following probability density function for a single random variable, Y:

$$f(y|\alpha,\beta) = \alpha\beta^{\alpha}y^{-\alpha-1}$$

where $\alpha > 0$, $\beta > 0$, $y \geq \beta$, and β is assumed known.

Consider the following statements:

I. The log-likelihood of $f(y)$ can be written as:

$$l(\alpha; y) = \ln\alpha + \alpha\ln\beta - (\alpha+1)\ln y$$

II. The score function, $U(\alpha)$, is:

$$U(\alpha) = \ln\left(\frac{\beta}{y}\right) + \alpha^{-1}$$

III. The Fisher Information, $I(\alpha)$, is:

$$I(\alpha) = \alpha^{-2}$$

Determine which of the above statements are true.

14% A I only

3% B II only

3% C III only

71% D I, II, and III

9% E The answer is not given by (A), (B), (C) or (D).

1/1 68% 3.9 3.9 5:30 4:00

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

You are given the following information for a fitted GLM:

Response Variable	Occurrence of Accidents
Response Distribution	Binomial
Link	Logit

Parameter	df	$\hat{\beta}$
Intercept	1	x
Driver's Age	2	
1	1	0.288
2	1	0.064
3	0	0
Area	2	
A	1	-0.036
B	1	0.053
C	0	0
Vehicle Body	2	
Bus	1	1.136
Other	1	-0.371
Sedan	0	0

The probability of a driver in age group 2, from area C and with vehicle body type Other, having an accident is 0.22.

Calculate the odds ratio of the driver in age group 3, from area C and with vehicle body type Sedan having an accident.

3% A Less than 0.200

5% B At least 0.200, but less than 0.250

20% C At least 0.250, but less than 0.300

4% D At least 0.300, but less than 0.350

68% E At least 0.350

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29

You are given the following information for a fitted GLM used to model policy renewals:

Response variable	Renewal outcome (1 or 0)
Response distribution	Bernoulli
Link	Logit

Observation i	Leverage, h_i	Deviance residual, d_i
1	?	0.214
2	0.049	0.218
3	0.043	0.221
⋮	⋮	⋮

Out of the first three observations, observation 1 has the largest standardized deviance residual.

Calculate the minimum possible value of h_1.

3% A Less than 0.035

13% B At least 0.035, but less than 0.055

14% C At least 0.055, but less than 0.075

11% D At least 0.075, but less than 0.095

60% At least 0.095

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

You are fitting a Poisson regression model of the form:

$$\mathrm{E}[Y_i] = \beta_1 + \beta_2 x_i$$

Maximum likelihood estimates of the beta coefficients are obtained using iterative weighted least squares procedure. You are given three matrixes:

- W is the weight matrix.
- X is the design matrix.
- Z has the beta values from the prior iteration applied to the explanatory variables as well as the correction term.

From the estimates of the first iteration, the following matrices are calculated using the matrixes as defined above:

$$(\mathrm{X^TWX})^{(1)} = \begin{bmatrix} 2.452 & -0.540 \\ -0.540 & 0.740 \end{bmatrix}$$

$$\left[(\mathrm{X^TWX})^{(1)}\right]^{-1} = \begin{bmatrix} 0.486 & 0.355 \\ 0.355 & 1.610 \end{bmatrix}$$

$$(\mathrm{X^TWz})^{(1)} = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$$

Calculate $b_2^{(2)}$, the estimate for β_2 on the second iteration.

1% A Less than 0

76% ✔ At least 0, but less than 0.40

17% C At least 0.40, but less than 0.80

6% D At least 0.80, but less than 1.20

1% E At least 1.20

0/1 61% 3.10 6.1 16:20 5:52

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

You are given the following information about three candidates for a Poisson frequency GLM on a group of condominium policies:

Model	Variables in the Model	DF	Log Likelihood	AIC	BIC
1	Risk Class	5	-47,704	95,418	95,473.61182
2	Risk Class + Region		-47,975		
3	Risk Class + Region + Claim Indicator	10	-47,365	94,750	

- Insureds are from one of five Risk Class: A, B, C, D, E.
- Condominium policies are located in several regions.
- Claim Indicator is either Yes or No.
- All models are built on the same data.

Calculate the absolute difference between the AIC and the BIC for Model 2.

Incorrect Answer

13% A Less than 85

20% B At least 85, but less than 95

61% C At least 95, but less than 105

3% D At least 105, but less than 115

3% E At least 115

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

You are given the following three statements regarding local regression:

I. Local regression is a memory-based procedure.

II. A small value of span s results in a global fit.

III. Local regression should not be used in a high-dimensional setting.

Determine which of the above statements are true.

Incorrect Answer

8% A None are true

5% B I and II only

66% I and III only

15% D II and III only

7% The answer is not given by (A), (B), (C) or (D).

1/1 61% 3.11 6.4 7:04 5:46

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

You have fit the following cubic spline model to a data set with 30 points:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 (x_i - \xi)_+^3 + \varepsilon_i$$

with

- $\hat{\beta}_0 = -1$
- $\hat{\beta}_1 = 4$
- $\hat{\beta}_2 = 3$
- $\xi = 1.5$

From this model you have calculated the following predicted values:

x_i	$\hat{y}_i$
1	4
3	-12.625

Determine the predicted value for this model when $x_i = 4$.

29% A Less than -52

4% B At least -52, but less than -50

61% At least -50, but less than -48

2% D At least -48, but less than -46

4% E At least -46