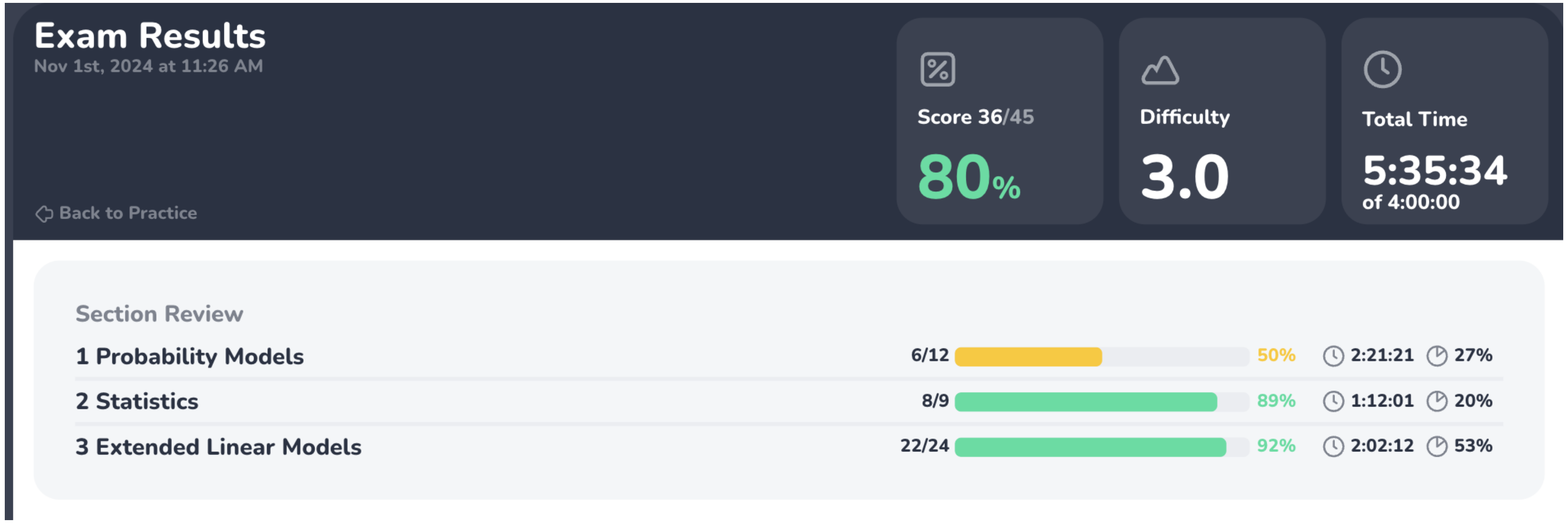
CA – Exam 1


Exam Results
Nov 1st, 2024 at 11:26 AM
Back to Practice
Score 36/45
80%
Difficulty
3.0
Total Time
5:35:34
of 4:00:00
Section Review
1 Probability Models
6/12
50%
2:21:21
27%
2 Statistics
8/9
89%
1:12:01
20%
3 Extended Linear Models
22/24
92%
2:02:12
53%

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

You are given the following information:

- The lifetimes of all light bulbs follow the exponential distribution.
- A new incandescent light bulb has a hazard rate of 0.20 per year.
- The expected lifetime of an LED light bulb is twice the expected future lifetime of a 2-year-old incandescent light bulb.

Calculate the hazard rate per year for an LED light bulb.

2% A Less than 0.05

5% B At least 0.05, but less than 0.07

11% C At least 0.07, but less than 0.09

74% D At least 0.09, but less than 0.11

8% E At least 0.11

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

The probability density function of X is:

$$f_X(x) = \frac{18}{(x+3)^3}, \quad x > 0$$

Let $Y = \sqrt{X}$.

Calculate the mode of Y.

Incorrect Answer

63% Less than 0.78

10% B At least 0.78, but less than 0.80

15% At least 0.80, but less than 0.82

5% D At least 0.82, but less than 0.84

7% E At least 0.84

0/1 62% 1.1 4.8 6:46 3:38

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

You are given the following information about a random variable, X.

- For all $r, s \geq 0, \Pr(X > r + s) = \Pr(X > r)\Pr(X > s)$.
- $\mathrm{E}[X \mid X > 10] = 30$

Calculate the result for the expression: $\mathrm{E}[X \mid X > 20]$.

Incorrect Answer

6% A Less than 30

20% At least 30, but less than 40

62% At least 40, but less than 50

4% D At least 50, but less than 60

7% E At least 60

1/1 65% 1.2 4.6 8:29 4:17

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

The distribution for loss amounts, X, is modeled as a mixture of three exponential distributions with the following means and weights:

Mean	Weight
400	0.5
600	0.4
1,000	0.1

An insurance coverage on the losses has a policy limit of 1,500.

Calculate the expected insurance payment size.

3% A Less than 450

65% At least 450, but less than 500

28% C At least 500, but less than 550

3% D At least 550, but less than 600

1% E At least 600

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

Steve catches fish at a Poisson rate of 3 per hour. The price Steve gets at the market for each fish is randomly distributed as follows:

Price	Probability
$10	20%
$20	60%
$30	20%

Using the normal approximation without a continuity correction, calculate the probability that Steve will receive at least $300 for fish caught in a four hour-period.

Incorrect Answer

6% A Less than 0.19

5% At least 0.19, but less than 0.20

77% At least 0.20, but less than 0.21

5% D At least 0.21, but less than 0.22

6% E At least 0.22

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

You are given the following information about a system:

- This is a 2-out-of-3 system.
- All components are independent.
- The probability of each component functioning is $p = 0.90$.

Calculate the reliability of the system.

4% A Less than 0.970

89% At least 0.970, but less than 0.975

2% C At least 0.975, but less than 0.980

1% D At least 0.980, but less than 0.985

4% E At least 0.985

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

You are given the following information about a system of three components:

- The minimal path sets are $A_1 = \{1, 2\}$ and $A_2 = \{2, 3\}$.
- All components in the system are independent.
- Components 1 and 3 have a reliability of 0.7.
- Component 2 has a reliability of 0.9.

What is the reliability of the system?

Incorrect Answer

4% A Less than 0.81

71% At least 0.81, but less than 0.82

3% C At least 0.82, but less than 0.83

1% D At least 0.83, but less than 0.84

21% At least 0.84

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

You are given the following information about a Markov chain with a transition probability matrix:

- $P = \begin{bmatrix} 0.60 & 0.30 & 0.10 \\ 0.70 & 0.30 & 0.00 \\ 0.30 & 0.00 & 0.70 \end{bmatrix}$
- The three states are 0, 1, and 2.

Calculate the long-run proportion of time in State 2.

Incorrect Answer

67% Less than 0.200

2% B At least 0.200, but less than 0.220

2% C At least 0.220, but less than 0.240

25% At least 0.240, but less than 0.260

4% E At least 0.260

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

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

- Planes move among three states servicing flights:
    - State 0: On-time
    - State 1: Delayed
    - State 2: Canceled
- $P = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.7 & 0.0 & 0.3 \\ 0.8 & 0.2 & 0.0 \end{bmatrix}$
- A plane is currently in the state Canceled.

Calculate the probability that the plane will be in the state Canceled after two transitions.

1% A Less than 0.07

2% B At least 0.07, but less than 0.09

13% C At least 0.09, but less than 0.11

4% D At least 0.11, but less than 0.13

At least 0.13

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

A particular temperature control system is modeled by a three state Markov chain, with states Cool (state 0), Warm (state 1), and Hot (state 2); the movement of the system through the states is governed by the following daily transition probability matrix:

$$P = \begin{bmatrix} 0.0 & 1.0 & 0.0 \\ 0.4 & 0.0 & 0.6 \\ 0.0 & 1.0 & 0.0 \end{bmatrix}$$

If the system is Cool on day 0, calculate the probability that it will be Hot on day 92.

8% A Less than 0.3

24% B At least 0.3, but less than 0.5

66% ✔ At least 0.5, but less than 0.7

1% D At least 0.7, but less than 0.9

1% E At least 0.9

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.

Incorrect Answer

14% A Less than 0.075

6% At least 0.075, but less than 0.085

65% At least 0.085, but less than 0.095

6% D At least 0.095, but less than 0.105

9% E At least 0.105

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

It is given that losses from an insurance product follow a Weibull distribution with $\tau = 3$ and $\theta = 5{,}000$. You want to simulate loss amounts using the inverse transformation method.

The uniform random numbers are simulated using a mixed congruential generator

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

with the following parameters:

$$\begin{aligned} x_0 &= 400 && \text{(the initial seed)} \\ a &= 61 && \text{(the multiplicative factor)} \\ c &= 593 && \text{(the additive term)} \\ m &= 1{,}024 && \text{(the modulus)} \end{aligned}$$

Calculate the average loss using three iterations of the generator.

5% A Less than 3,000

7% B At least 3,000, but less than 3,250

8% C At least 3,250, but less than 3,500

73% At least 3,500, but less than 3,750

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

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

$$\sum_{i=1}^{10} X_i = 100 \qquad \sum_{i=1}^{10} X_i^2 = 2{,}000$$

The parameters of this distribution are estimated using the method of moments with raw moments only.

Calculate the estimated variance of this distribution.

88% — A. Less than 120

3% — B. At least 120, but less than 140

3% — C. At least 140, but less than 160

1% — D. At least 160, but less than 180

5% — E. At least 180

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

You are given:

- A random variable X has an exponential distribution with mean θ.
- An estimator of θ has a mean of 2 and a variance of 0.5.

Calculate the mean square error of this estimator for $\theta = 2.1$.

6% A Less than 0.3

3% B At least 0.3, but less than 0.4

8% C At least 0.4, but less than 0.5

80% At least 0.5, but less than 0.6

2% E At least 0.6

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

You are given the following information:

- A random variable X follows a gamma distribution with parameters $\alpha = 0.6$ and unknown θ.
- Null hypothesis is $H_0 : \theta = 2{,}000$
- Alternative hypothesis is $H_1 : \theta > 2{,}000$
- n values of X will be observed and the null hypothesis is rejected if:

$$\frac{1}{n}\sum_{i=1}^{n} X_i > 1{,}500$$

Calculate the minimum value of n which will result in a probability of Type I error of less than 2.5%, using the normal approximation.

10% A Less than 100

67% At least 100, but less than 105

13% C At least 105, but less than 110

5% D At least 110, but less than 115

4% E At least 115

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

Given

- $X_1, X_2, \ldots, X_{10}$ are random samples from a Normal distribution with mean μ_x and variance 7.
- $Y_1, Y_2, \ldots, Y_8$ are random samples from a Normal distribution with mean μ_y and variance 9.
- X and Y are independent.
- $W = \bar{X} - \bar{Y}$ is a Normal distribution with mean $\mu_x - \mu_y$.
- $H_0 : \mu_x = \mu_y$
- $H_1 : \mu_x > \mu_y$

For $W = 2$, what is the result of the hypothesis test using the Normal distribution?

4% A Reject H_0 at the 0.025 level

7% B Reject H_0 at the 0.050 level, but not at the 0.025 level

68% Reject H_0 at the 0.075 level, but not at the 0.050 level

10% D Reject H_0 at the 0.100 level, but not at the 0.075 level

11% E Do not reject H_0 at the 0.100 level

0/1 67% 2.4 4.7 9:07 5:39

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

You are given the following information:

- A random variable follows the normal distribution with mean μ and standard deviation 3.
- H_0: $\mu = 8$
- H_1: $\mu = 7$
- It is required that errors of Type I and II have probabilities of 0.05 and 0.01, respectively.

Calculate the necessary sample size based on the Neyman-Pearson Lemma.

Incorrect Answer

10% A Less than 50

13% At least 50, but less than 100

67% At least 100, but less than 150

8% D At least 150, but less than 200

2% E At least 200

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

You are testing whether the number of claims in year 2 is independent of the number of claims in year 1.

You are given the following contingency table:

		Claims in Year 2	
		0	1+
Claims in Year 1	0	20	21
	1+	30	21

Calculate the Chi-Square test statistic.

4% A Less than 0.5

78% B At least 0.5, but less than 1.0

8% C At least 1.0, but less than 1.5

4% D At least 1.5, but less than 2.0

6% E At least 2.0

1/1 70% 2.6 4.4 5:00 4:13

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

You are given the following two samples:

$$\text{Sample 1}: \quad 14 \quad 15 \quad 17 \quad 20 \quad 27 \quad 29$$

$$\text{Sample 2}: \quad 27 \quad 32 \quad 34 \quad 35 \quad 36$$

These samples are taken from two independent normal distributions with means μ_1 and μ_2, and variances 36 and 64, respectively.

Calculate the upper bound of the 95% symmetric confidence interval for the difference $\mu_2 - \mu_1$.

14% A Less than 20

70% At least 20, but less than 22

7% C At least 22, but less than 24

4% D At least 24, but less than 26

4% E At least 26

1/1 | 75% | 2.7 | 2.9 | 12:57 | 3:13

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

Let $Y_1 < Y_2 < \cdots < Y_{15}$ be the order statistics of a random sample from a uniform distribution on $[0, 1]$.

Calculate the probability that $0.75 < Y_{15} < 0.85$.

6% A Less than 5%

3% B At least 5%, but less than 6%

6% C At least 6%, but less than 7%

75% D At least 7%, but less than 8%

10% E At least 8%

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

You are given the following information:

- For a general liability policy loss amounts, Y, follow the Pareto distribution with probability density function:

$$f(y) = 3\theta^3 (y+\theta)^{-4}, \quad \theta = 1{,}000, \quad y > 0$$

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

Calculate the 90th percentile of Y_{10}.

7% A. Less than 3,500

8% B. At least 3,500, but less than 3,550

74% C. At least 3,550, but less than 3,600

7% D. At least 3,600, but less than 3,650

3% E. At least 3,650

1/1 69% 3.1 4.4 2:52 1:26

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

You are given the following statements on the bias-variance tradeoff:

I. Bias refers to the error arising from the method's sensitivity towards the training data set.
II. Variance refers to the error arising from the assumptions made in the statistical learning tool.
III. The variance of a statistical learning method increases as the method's flexibility increases.

Determine which of the statements is/are true.

5% A I only

2% B II only

69% III only

19% D I, II, and III

4% 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

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.

3% A Less than -2

72% At least -2, but less than 0

20% C At least 0, but less than 2

3% D At least 2, but less than 4

2% E At least 4

1/1 81% 3.3 1.2 14:16 3:59

14
15
16
17
18
19
20
21
22
23
24
25
26

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

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

Calculate the adjusted R^2 for this model.

2% A Less than 0.70

5% B At least 0.70, but less than 0.72

81% At least 0.72, but less than 0.74

9% D At least 0.74, but less than 0.76

2% E At least 0.76

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

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

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

His analysis results in the following output:

Coefficients	Estimate	Std. Error	t-value	$\Pr(>\lvert t\rvert)$
Intercept	18,169.300	212.2080	85.62027	2.05E-14
Experience	1,110.233	59.8224	18.55881	1.75E-08
Gender	169.550	162.9177	10.38285	2.62E-06

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

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

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

15% A At most 18,169.3

74% Greater than 18,169.3, but at most 18,400.0

4% C Greater than 18,400.0, but at most 18,600.0

1% D Greater than 18,600.0

5% E The answer cannot be computed from the information given.

1/1 80% 3.4 1.5 5:55 2:43

10
11
12
13
14
15
16
17
18
19
20
21
22

You are given the following information on a one-factor ANOVA regression:

Source	Degrees of Freedom	Sum of Squares
Mean	1	50,147
Treatments	3	5,028
Residuals	20	11,279
Total	24	66,454

At which significance level do we reject the hypothesis that the factor is not significant?

8% A Do not reject at 10% significance

80% Rejected at 10% significance, but not at 5% significance

7% C Rejected at 5% significance, but not at 2% significance

1% D Rejected at 2% significance, but not at 1% significance

4% E Rejected at 1% significance

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

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")
- The partial ANOVA table is given below:

Response variable	Losses
Response distribution	Normal
Link	Identity

Source of variation	Degrees of freedom	Sum of squares	Mean square
Number of Drivers	2	34,958.65	
Territory	4		6,452.60
Residuals		67,940.00	8,492.50

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

Calculate the p-value of this test.

2% A Less than 2%

3% B At least 2%, but less than 5%

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

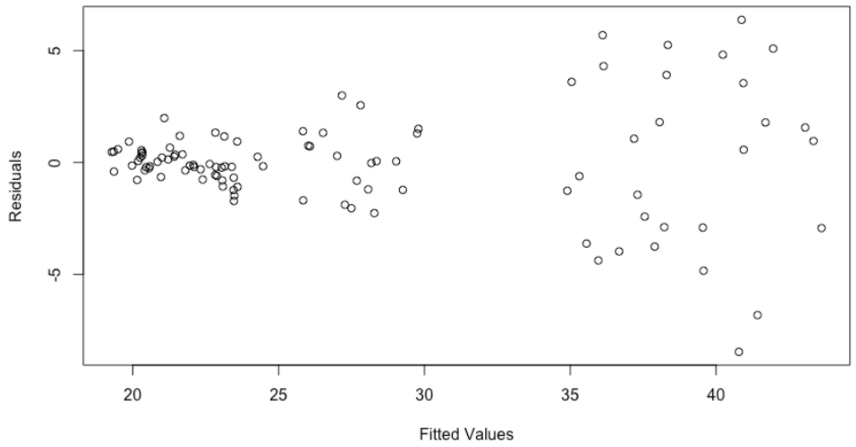
78% D At least 10%, but less than 20%

12% E At least 20%

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

You are given the following residual plot from a linear regression model fitted to a dataset:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$



Which of the following statements best describes the fitted model?

1% A The fitted model is appropriate for the dataset.

2% B There are high leverage points that should be removed from the dataset.

6% C The plot shows sign of multicollinearity, which can be addressed by removing predictors with high variance inflation factors.

82% ✔ The fitted model is not appropriate as the residual plot exhibits signs of heteroscedasticity.

9% E A square transformation of the response variable should be performed to address the issue of heteroscedasticity.

1/1 | 78% | 3.5 | 2.3 | 2:55 | 1:22

9
10
11
12
13
14
15
16
17
18
19
20
21
22

Determine which of the following statements is true regarding leverage.

5% A Its smallest possible value is 0.

2% B It is a function of the response variable.

8% C A large value indicates the presence of an outlier.

7% D They are the diagonal elements of $(\mathbf{X}^T\mathbf{X})^{-1}$ where $\mathbf{X}$ is the design matrix.

78% They are the diagonal elements of $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ where $\mathbf{X}$ is the design matrix.

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

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

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

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

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

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

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

4% A Less than 1,000

7% B At least 1,000, but less than 1,500

77% C At least 1,500, but less than 2,000

2% D At least 2,000, but less than 2,500

9% E At least 2,500

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

Best subset selection is performed on a data set with three predictor variables: X_1, X_2, and X_3.

The table below displays the R^2 and adjusted R^2 for all possible linear regression models.

Model	Predictors	R^2	Adjusted R^2
I	None	-	-
II	X_1	0.6215	0.6054
III	X_2	0.5186	0.4981
IV	X_3	0.4539	0.4307
V	X_1, X_2	0.8136	0.8014
VI	X_1, X_3	0.7236	0.7056
VII	X_2, X_3	0.8150	0.8029
VIII	X_1, X_2, X_3	0.8160	0.7996

Determine the model that is selected as a result of best subset selection.

8% A Model V

1% B Model VI

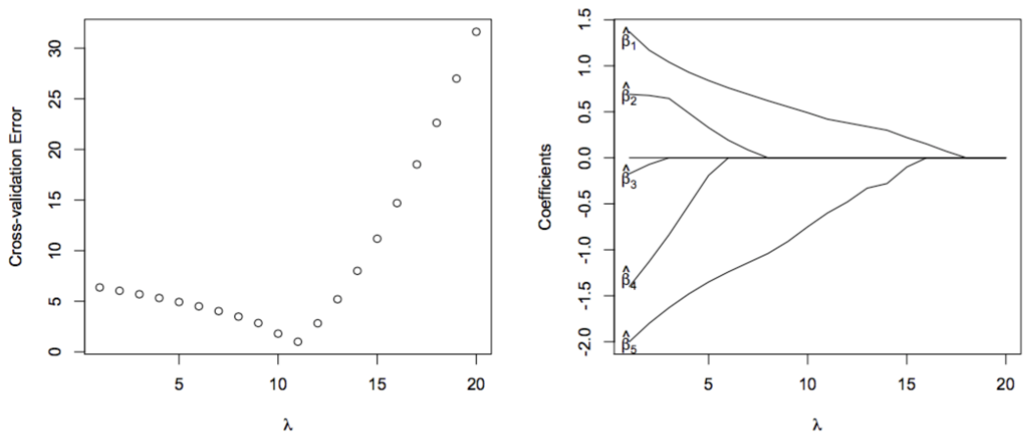
81% Model VII

9% D Model VIII

1% E The correct 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
18
19

You use lasso to fit a linear regression model with five potential continuous predictors, X_1 through X_5. After fitting the model, you use cross-validation to choose the best model, producing the following cross-validation and coefficient plots:



Note that the subscript of the coefficient corresponds to the predictor.

Determine which of the following predictors are chosen by this procedure. (select all that apply)

78% X_1

8% B X_2

7% C X_3

7% D X_4

77% X_5

1/1 72% 3.7 3.5 14:08 5:28

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

You are fitting the linear regression model

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

to the following set of 4 data points. The regression is performed without standardizing the variables.

Obs.	1	2	3	4
y_i	1.1	5.4	14.7	25.7
$x_{i,1}$	3	4	4	7
$x_{i,2}$	10	9	7	5

Two of the models under consideration are:

- Model I: $\hat{\beta}_0 = 29.88$, $\hat{\beta}_1 = 1.62$, $\hat{\beta}_2 = -3.29$
- Model II: $\hat{\beta}_0 = 37.74$, $\hat{\beta}_1 = 1.12$, $\hat{\beta}_2 = -4.00$

One of the potential models is the result of ridge regression with tuning parameter $\lambda = 5$.

Calculate the predicted value of Y under that model when $X_1 = 5$ and $X_2 = 8$.

2% A Less than 11.3

22% B At least 11.3, but less than 11.5

72% C At least 11.5, but less than 11.7

3% D At least 11.7, but less than 11.9

1% E At least 11.9

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

Wahlberg is investigating the performance of weightlifters in a tournament based on a set of variables. He collects the following information on 100 weightlifters:

- X_1 : Body mass index, BMI
- X_2 : Body fat percentage, BFP (0 to 100)

The following summary statistics for the two variables are given:

- $\sum_{i=1}^{100} x_{i,1} = 2{,}623$
- $\sum_{i=1}^{100} x_{i,1}^2 = 71{,}549$
- $\sum_{i=1}^{100} x_{i,2} = 972$
- $\sum_{i=1}^{100} x_{i,2}^2 = 9{,}946$

A principal component analysis is performed on the data set without scaling the variables. The results of the principal component analysis are as follows:

- The first principal component loading for the first variable is 0.6.
- The first principal component loading for the second variable is negative.

Calculate the first principal component score for a weightlifter with a BMI of 27.0 and a BFP of 11.0.

2% A Less than -1.0

71% At least -1.0, but less than -0.5

15% C At least -0.5, but less than 0.0

5% D At least 0.0, but less than 0.5

7% E At least 0.5

1/1 | 76% | 3.8 | 3.2 | 4:44 | 1:00

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

Determine which of the following distribution and link function should be used for slices of pizzas sold at a convenient store based on distance to the city center.

11% A Normal distribution, identity link function

2% B Normal distribution, log link function

8% C Poisson distribution, identity link function

76% ✔ Poisson distribution, log link function

3% E Inverse Gaussian distribution, log link function

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

You are given the outputs from two GLMs fitted to the same data from a trial of a new drug.

Model 1

Response variable	Number
Response distribution	Poisson
Link	Log
AIC	273.877

Parameter	$\hat{\beta}$	$s.e.\left(\hat{\beta}\right)$
Intercept	4.529	0.147
Treatdrug		
Placebo	0.000	0.000
Drug	-1.359	0.118
Age	-0.039	0.006

Model 2

Response variable	Number
Response distribution	Negative binomial
Link	Log
AIC	164.880

Parameter	$\hat{\beta}$	$s.e.\left(\hat{\beta}\right)$
Intercept	4.526	0.595
Treatdrug		
Placebo	0.000	0.000
Drug	-1.368	0.369
Age	-0.039	0.021

Determine which of the following statements is false using the Wald test.

5% A Under Model 1, the *TreatDrug* coefficient has a p-value less than 0.01.

9% B Under Model 1, the *Age* coefficient has a p-value less than 0.01.

11% C Under Model 2, the *TreatDrug* coefficient has a p-value less than 0.01.

67% ✔ Under Model 2, the *Age* coefficient has a p-value less than 0.01.

7% E Under both models, the intercept coefficient has a p-value less than 0.01.

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

Given a family of distributions where the variance is related to the mean through a power function:

$$\mathrm{Var}\left[Y\right] = a\mathrm{E}\left[Y\right]^p$$

One can characterize members of the exponential family of distributions using this formula.

You are given the following statements on the value of p for a given distribution:

I. Normal (Gaussian) distribution, $p = 0$
II. Compound Poisson-gamma distribution, $1 < p < 2$
III. Inverse Gaussian distribution, $p = -1$

Determine which of the above statements are correct.

Incorrect Answer

8% A I only

71% I and II only

8% C I and III only

8% D II and III only

5% The answer is not given by (A), (B), (C), or (D)

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

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 $\hat{\pi}$.

7% A Less than 0.60

16% B At least 0.60, but less than 0.70

74% At least 0.70, but less than 0.80

2% D At least 0.80, but less than 0.90

1% E At least 0.90

10
11
12
13
14
15
16
17
18
19
20
21
22

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.

2% A Less than 20%

74% ✔ At least 20%, but less than 40%

19% C At least 40%, but less than 60%

4% D At least 60%, but less than 80%

1% E At least 80%

1/1 | 78% | 3.9 | 1.9 | 2:51 | 2:42

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

You are given the following information for two potential logistic models used to predict the occurrence of a claim:

- Model 1: AIC = 262.68

Parameter	$\hat{\beta}$
(Intercept)	-3.264
Vehicle Value ($000s)	0.212
Gender - Female	0.000
Gender - Male	0.727

- Model 2: AIC = 263.39

Parameter	$\hat{\beta}$
(Intercept)	-2.894
Gender - Female	0.000
Gender - Male	0.727

- AIC is used to select the most appropriate model.

Calculate the probability of a claim for a male policyholder with a vehicle valued $12,000 by using the selected model.

16% A Less than 0.15

1% B At least 0.15, but less than 0.30

3% C At least 0.30, but less than 0.45

78% D At least 0.45, but less than 0.60

2% E At least 0.60

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

You are given the following information for a fitted GLM:

Response Variable	Y
Response Distribution	Poisson
Link	Log
AIC	221.254

Parameter	$\hat{\beta}$	s. e. $(\hat{\beta})$
Intercept	5.421	0.228
Gender		
Male	0.000	0.000
Female	-0.557	0.217
Age	0.107	0.002

Calculate the predicted value of Y for a Female with an Age value of 22.

1% A Less than 1,000

96% At least 1,000, but less than 1,500

3% C At least 1,500, but less than 2,000

1% D At least 2,000, but less than 2,500

0% E At least 2,500

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 local regression model

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

to the following set of 20 data points using a span of $s = 0.20$.

Obs.	1	2	3	4	5	6	7	8	9	10
y_i	2.7	3.7	6.4	11.9	16.4	15.4	15.8	16.1	17.0	19.2
x_i	2.2	4.8	6.9	9.2	10.2	10.8	12.0	13.4	13.8	14.1

Obs.	11	12	13	14	15	16	17	18	19	20
y_i	20.1	21.3	27.9	25.9	23.9	25.5	28.2	29.1	26.7	27.1
x_i	15.0	15.4	15.8	16.4	17.4	19.0	19.2	19.4	22.8	26.5

When fitting the linear local regression at $X = 14$, you are using the weighting function:

$$K_i = 1 - |14 - x_i|$$

This linear local regression at $X = 14$ has produced the estimate $\hat{\beta}_0 = -47.21$.

Calculate the predicted value of Y when $X = 14$.

12% A Less than 18.35

67% At least 18.35, but less than 18.55

12% C At least 18.55, but less than 18.75

5% D At least 18.75, but less than 18.95

4% E At least 18.95

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 local regression model

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

to the following set of 20 data points using a span of $s = 0.2$:

Obs.	1	2	3	4	5	6	7	8	9	10
y_i	2.7	3.7	6.4	11.9	16.4	15.4	15.8	16.1	17.0	19.2
x_i	2.2	4.8	6.9	9.2	10.2	10.8	12.0	13.8	13.9	14.1

Obs.	11	12	13	14	15	16	17	18	19	20
y_i	20.1	21.3	27.9	25.9	23.9	25.5	28.2	29.1	26.7	27.1
x_i	14.2	14.4	15.8	16.4	17.4	19.0	19.2	19.4	22.8	26.5

When fitting the linear local regression at $X = 17$, you are using the weighting function

$$K_i = \frac{2 - |17 - x_i|}{2}$$

This linear local regression at $X = 17$ has produced the estimate $\hat{\beta}_0 = 65$.

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

6% A Less than -3

72% ✔ At least -3, but less than -2

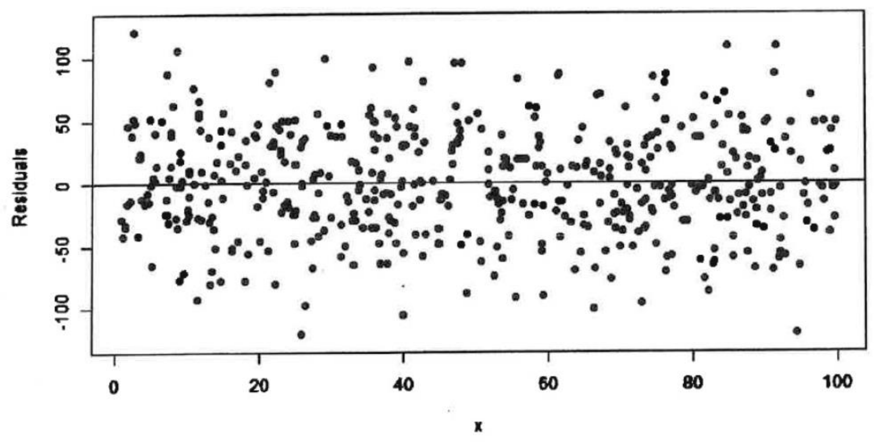
14% C At least -2, but less than -1

5% D At least -1, but less than 0

2% E At least 0

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

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



Determine which of the following enhancements to the model would most likely improve the fit to the testing data set.

8% A Linear Spline

12% B Local Regression

7% C Polynomial Regression

2% D Step Function

70% There is no evidence that any of (A), (B), (C), (D) will improve the fit.

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 constant local regression model

$$Y = \beta_0 + \varepsilon$$

to the following set of 20 data points using a span of $s = 0.3$:

Obs.	1	2	3	4	5	6	7	8	9	10
y_i	2.7	3.7	6.4	11.9	16.4	15.4	15.8	16.1	17.0	19.2
x_i	2.2	4.8	6.9	9.2	10.2	10.8	12.0	13.8	14.0	14.2

Obs.	11	12	13	14	15	16	17	18	19	20
y_i	20.1	21.3	27.9	25.9	23.9	25.5	28.2	29.1	26.7	27.1
x_i	15.2	15.4	15.8	16.4	17.4	19.0	19.2	19.4	22.8	26.5

When fitting the constant local regression at $X = 15$, you are using the weighting function

$$K_i = 1.44 - (15 - x_i)^2$$

Calculate the fitted value at $X = 15$.

Incorrect Answer

7% A Less than 18

6% At least 18, but less than 19

13% C At least 19, but less than 20

10% D At least 20, but less than 21

63% At least 21