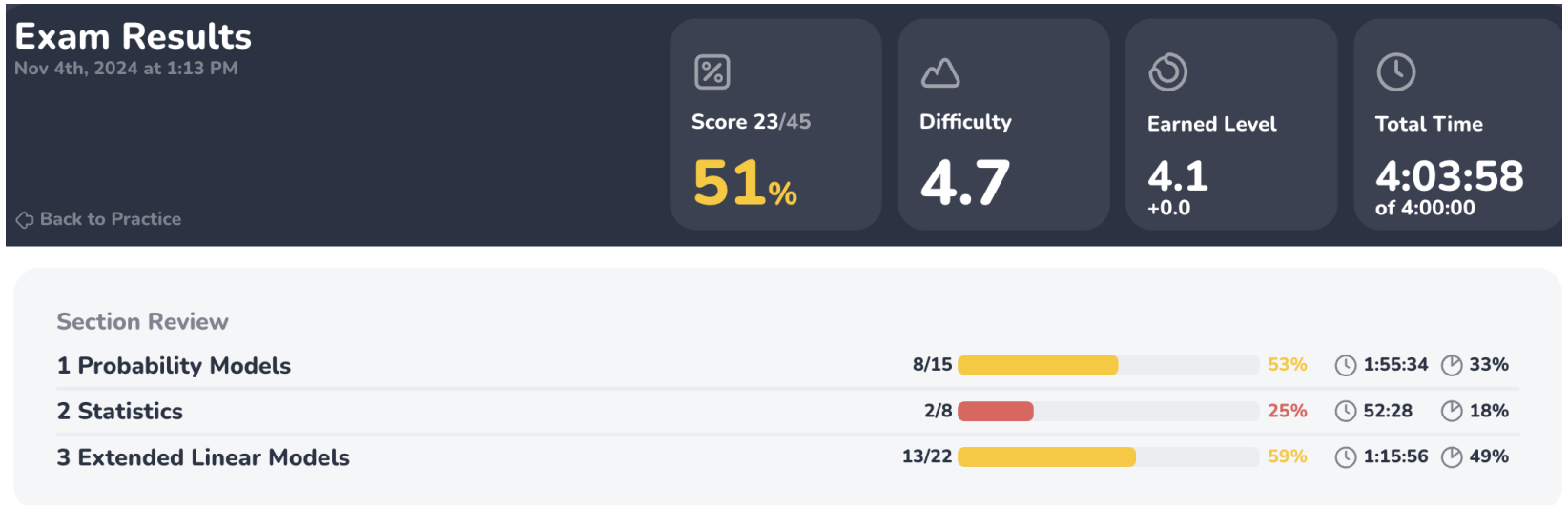
CA – Exam 3


Exam Results
Nov 4th, 2024 at 1:13 PM
Back to Practice
Score 23/45
51%
Difficulty
4.7
Earned Level
4.1
+0.0
Total Time
4:03:58
of 4:00:00
Section Review
1 Probability Models
8/15
53%
1:55:34
33%
2 Statistics
2/8
25%
52:28
18%
3 Extended Linear Models
13/22
59%
1:15:56
49%

You are given the following information:

- The amount of time one spends in an IRS office is exponentially distributed with a mean of 30 minutes.
- P_1 is the probability that John will spend more than an hour of total time in the IRS office, given that he has already been in the office for 20 minutes.
- P_2 is the probability that Lucy will spend more than an hour of total time in the IRS office, given that she has just arrived.

Calculate the difference $(P_1 - P_2)$.

8% A Less than 0.00

16% B At least 0.00, but less than 0.05

2% C At least 0.05, but less than 0.10

72% ✓ At least 0.10, but less than 0.15

2% E At least 0.15

You are given the following information:

- At the time of installation a given machine has the following survival function:

$$S(x) = \left(1 - \frac{x}{100}\right)^{1/2}, \quad \text{for } 0 \leq x \leq 100, \quad x \text{ in months}$$

- After 16 months, that machine is still functioning.

Calculate the probability that the machine will stop working between months 36 and 51.

5% A Less than 0.095

15% B At least 0.095, but less than 0.105

73% ✓ At least 0.105, but less than 0.115

2% D At least 0.115, but less than 0.125

5% E At least 0.125

You are given the probability density function for a random variable X:

$$f_X(x) = \left(\frac{\lambda}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\lambda-1} \exp\left[-\left(\frac{x}{\alpha}\right)^{\lambda}\right], \quad x \geq 0, \quad \lambda > 0, \quad \alpha > 0$$

Determine which of the following statements is/are true.

I. If $\alpha = 1$, then X follows a standard Weibull distribution.
II. If $\lambda = 1$, then X follows an exponential distribution with mean α^{-1}.
III. If $\alpha = \lambda = 2$, the mean of X is $3\sqrt{\pi}$.

61% ✔ I only

10% B II only

7% C III only

8% D I, II, and III

13% E The correct answer is not given by (A), (B), (C), or (D).

A random variable, X, is uniformly distributed on the interval (m, n).

Calculate the CTE_q of X.

Incorrect Answer

5% A $\frac{q\,(n+m)}{2}$

9% $\frac{q\,(n-m)}{2}$

15% C $\frac{(1-q)\,(n+m)}{2}$

58% $\frac{n\,(1+q)+m\,(1-q)}{2}$

13% E $\frac{n\,(1-q)+m\,(1+q)}{2}$

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

You are given:

- Company CA has one representative who answers forum questions.
- Forum questions are posted at a Poisson rate of 10 per hour.
- The time taken for the representative to answer each question is exponentially distributed with a mean of 10 minutes.
- The times questions are posted and the time taken to answer them are independent.
- There are at least two unanswered questions in the system.

Calculate the probability of the representative answering two questions before four new questions are posted.

11% A Less than 0.60

58% ✓ At least 0.60, but less than 0.62

14% C At least 0.62, but less than 0.64

6% D At least 0.64, but less than 0.66

11% E At least 0.66

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

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

- The minimal cut sets are $C_1 = \{1, 4\}$ and $C_2 = \{2, 3\}$.
- All components in the system are independent.
- Components 1 and 3 have a reliability of 0.90.
- Components 2 and 4 have a reliability of 0.70.

What is the reliability of the system?

✖ Incorrect Answer

17% A Less than 0.93

4% ✖ At least 0.93, but less than 0.94

71% ✔ At least 0.94, but less than 0.95

2% D At least 0.95, but less than 0.96

6% E At least 0.96

You are given the following information:

- Consider a system of three independent components, each of which functions for an amount of time (in months) uniformly distributed over (0, 1).
- Under current design, the system will fail if any of the components fail.
- A new design was made such that the system will fail if 2 or more components fail.

Calculate the increase in expected system life in months under the new system design.

8% A Less than 0.195

5% B At least 0.195, but less than 0.215

10% C At least 0.215, but less than 0.235

65% ✓ At least 0.235, but less than 0.255

11% E At least 0.255

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

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

- Two out of three components are required for the system to function.
- Each component has a lifetime that is exponential with mean 1.

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

✖ Incorrect Answer

3% A $\int_0^\infty e^{-3t}\,dt$

2% ✖ $\int_0^\infty te^{-3t}\,dt$

11% C $\int_0^\infty \left(e^{-2t}-e^{-3t}\right)dt$

72% ✔ $\int_0^\infty \left(3e^{-2t}-2e^{-3t}\right)dt$

12% E $\int_0^\infty \left(3te^{-2t}-2te^{-3t}\right)dt$

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

You are given the following Markov chain transition matrix (states 1, 2, 3, 4, 5):

$$\mathbf{P} = \begin{bmatrix} 0.1 & 0.9 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.5 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.8 & 0.2 \\ 0.0 & 0.0 & 0.6 & 0.4 & 0.0 \\ 0.3 & 0.0 & 0.6 & 0.0 & 0.1 \end{bmatrix}$$

You are also given the following statements:

1. There are two classes in this Markov chain.
2. State 1 is recurrent.
3. State 5 is transient.

Determine which of the following is/are correct.

✖ Incorrect Answer

5% A I only

8% B II only

8% C III only

68% ✔ I, II, and III

12% ✖ The correct answer is not given by (A), (B), (C), or (D).

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

- There are three daily wildfire risk states: Green (state 0), Yellow (state 1) and Red (state 2).
- Transition between states occurs at the end of each day.
- The daily transition matrix, $\mathbf{P} = \begin{bmatrix} 0.82 & m & n \\ 0.61 & 0.28 & 0.11 \\ 0.40 & 0.31 & 0.29 \end{bmatrix}$.
- The wildfire risk was Yellow on Wednesday.
- Today is Thursday and the wildfire risk is Green.
- The probability that the wildfire risk will be Red on Saturday is 0.07.

Calculate the absolute difference between *m* and *n*.

13% A Less than 0.065

6% B At least 0.065, but less than 0.075

69% ✓ At least 0.075, but less than 0.085

6% D At least 0.085, but less than 0.095

6% E At least 0.095

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

You are given:

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

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

✖ Incorrect Answer

8%	A	Less than 40%
7%	B	At least 40%, but less than 45%
12%	C	At least 45%, but less than 50%
62%	✔	At least 50%, but less than 55%
11%	✖	At least 55%

You are given the following information:

- A life insurance company issues a special 3-year discrete insurance to a life (x).
- If the policyholder dies, at the end of the year of death, there is a random drawing. With probability 0.2, the death benefit is 50,000. With probability 0.8, the death benefit is 0.
- At the beginning of each year the policy is in effect which (x) is alive, there is a random drawing. With probability 0.8, the premium π is paid. With probability 0.2, no premium is paid.
- The random drawings are independent.
- The probability of surviving an additional k years is ${}_kp_x = 0.9^k$, for $k = 0, 1, 2, \ldots$.
- $i = 0.06$

Calculate π using the equivalence principle.

✖ Incorrect Answer

10% A Less than 1,100

55% ✔ At least 1,100, but less than 1,200

10% ✖ At least 1,200, but less than 1,300

7% D At least 1,300, but less than 1,400

18% E At least 1,400

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

You are given the following information:

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

Calculate the actuarial present value of this annuity.

6% A Less than 10.0

7% B At least 10.0, but less than 10.5

12% C At least 10.5, but less than 11.0

10% D At least 11.0, but less than 11.5

64% ✓ At least 11.5

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

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

- X has density function: $f(x) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$, for $x > 0$.
- The rejection method is based on $g(x) = \exp(-x)$, for $x > 0$.
- The rejection procedure is as follows:
  - Step 1: Generate independent random numbers Y and U, where Y follows density function g and U is uniform on (0,1).
  - Step 2: If $U \leq \frac{f(Y)}{cg(Y)}$ stop and set $X = Y$. Otherwise return to Step 1.

Calculate the minimum possible value for c.

✖ Incorrect Answer

5%	A	Less than 0.50
6%	✖	At least 0.50, but less than 0.75
15%	C	At least 0.75, but less than 1.00
14%	D	At least 1.00, but less than 1.25
60%	✔	At least 1.25

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

You are given:

- X is a beta random variable with $a = 2, b = 3, \theta = 10$.
- X is simulated using the acceptance-rejection method with the following random variable with density function:

$$g(y) = \frac{1}{10}, \text{ where } 0 < y < 10$$

- The following values from the random variable above are given: $7.56, 2.36, 4.51, 8.27$.

The uniform random numbers are simulated using a mixed congruential generator

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

with the following parameters:

$$\begin{aligned} x_0 &= 300 && \text{(the initial seed)} \\ a &= 61 && \text{(the multiplicative factor)} \\ c &= 593 && \text{(the additive term)} \\ m &= 1{,}024 && \text{(the modulus)} \end{aligned}$$

Calculate the average of the simulated values of X.

6% A Less than 3

61% ✓ At least 3, but less than 4

18% C At least 4, but less than 5

11% D At least 5, but less than 6

5% E At least 6

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

You have a random sample of five observations:

$$2.5 \quad 7.5 \quad 12.5 \quad 16.0 \quad 17.5$$

- The probability density function below is fit to the random sample.

$$f(x) = \frac{2}{\beta^2} x_i e^{-\left(\frac{x_i}{\beta}\right)^2}$$

Calculate the maximum likelihood estimate of β.

✖ Incorrect Answer

9% A Less than 10

4% B At least 10, but less than 11

13% ✖ At least 11, but less than 12

66% ✔ At least 12, but less than 13

8% E At least 13

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

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:

$$f(x;\gamma,\theta) = \frac{\gamma x^{\gamma-1}}{\theta^{\gamma}\left[1+\left(\frac{x}{\theta}\right)^{\gamma}\right]^2}$$

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

Calculate the estimated value of θ.

✖ Incorrect Answer

3% A Less than 20

12% ✖ At least 20, but less than 30

20% C At least 30, but less than 40

8% D At least 40, but less than 50

57% ✔ At least 50

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

Determine the Fisher Information of n independent samples from a geometric distribution with mean β.

✖ Incorrect Answer

1% A $\frac{1}{\beta}$

5% ✖ $\frac{n}{\beta}$

72% ✔ $\frac{n}{\beta} - \frac{n}{\beta+1}$

9% D $\frac{n\beta}{\beta+1}$

13% E $\frac{\beta(\beta+1)}{n}$

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

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population X with probability mass function:

$$p(x) = \theta(1-\theta)^x \quad \text{for } x = 0, 1, 2, \ldots$$

θ is an unknown parameter between 0 and 1 and the expected value of X is:

$$\mathrm{E}(X) = \frac{1-\theta}{\theta}$$

Determine the Cramer-Rao lower bound for the variance of all unbiased estimators of θ.

✖ Incorrect Answer

14% A $\frac{1}{n}\theta(1-\theta)$

6% ✖ $n\theta(1-\theta)$

9% C $\frac{1}{n^2}\theta(1-\theta)$

64% ✔ $\frac{1}{n}\theta^2(1-\theta)$

6% E $\frac{1}{n^2}\theta^2(1-\theta)$

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

A random sample of 10 screw lengths is taken. The sample mean is 2.5 and the unbiased sample variance is 3.0. The underlying distribution is assumed to be normal.

You want to perform a test of the variance of screw lengths, σ^2.

- Null hypothesis is $H_0 : \sigma^2 = 6$
- Alternative hypothesis is $H_1 : \sigma^2 < 6$

Determine the result of this hypothesis test.

✖ Incorrect Answer

11% A Reject H_0 at the 0.005 level

6% ✖ Reject H_0 at the 0.010 level, but not at the 0.005 level

9% C Reject H_0 at the 0.025 level, but not at the 0.010 level

8% D Reject H_0 at the 0.050 level, but not at the 0.025 level

66% ✔ Do not reject H_0 at the 0.050 level

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

You are given an independent random sample from a normal distribution ***X*** with unknown mean, μ. You are given the following information:

- $\sigma^2(X) = 400$
- $H_0 : \mu = 0$
- $H_1 : \mu = 10$
- You will reject the null hypothesis if $\bar{X} > \alpha$ for some value of α.

Calculate the minimum sample size needed so that the probability of a Type I error and the probability of a Type II error are each no more than 10%.

9% A Less than 10

10% B At least 10, but less than 15

14% C At least 15, but less than 20

6% D At least 20, but less than 25

61% At least 25

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

You are given the following random observations:

$$0.1 \quad 0.2 \quad 0.5 \quad 1.0 \quad 1.3$$

You test whether the sample comes from a distribution with probability density function:

$$f(x) = \frac{2}{(1+x)^3}, \quad x > 0$$

Calculate the Kolmogorov-Smirnov statistic.

7% A Less than 0.18

78% At least 0.18, but less than 0.20

6% C At least 0.20, but less than 0.22

2% D At least 0.22, but less than 0.24

7% E At least 0.24

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

A particular model of car comes in three colors, Blue, Silver, and White. You want to test the null hypothesis that Blue cars are twice as popular as Silver cars, which are twice as popular as White cars.

You take a random sample of 100 cars and collect the following data:

Color	Count
Blue	46
Silver	32
White	22

Calculate the p-value for this hypothesis test.

✖ Incorrect Answer

6% ✖ Less than 0.005

11% B At least 0.005, but less than 0.010

13% C At least 0.010, but less than 0.025

61% ✔ At least 0.025, but less than 0.050

9% E At least 0.050

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

For any statistical learning method, which of the following increases monotonically as flexibility increases?

I. Training MSE
II. Test MSE
III. Bias squared
IV. Variance

✖ Incorrect Answer

4% A I and III

16% ✖ I and IV

3% C II and III

18% D II and IV

59% ✔ The correct answer is not given by (A), (B), (C), or (D).

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

You are given a random sample of four observations from a population:

$$\{1, 3, 10, 40\}$$

and would like to calculate the standard error of an estimate of the mean of the population using a bootstrap procedure. Your calculation is based on only two bootstrapped data sets.

Calculate the maximum possible value of your standard error estimate.

3% A Less than 10

22% B At least 10, but less than 20

60% ✓ At least 20, but less than 30

10% D At least 30, but less than 40

4% E At least 40

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

The model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ was fit using 6 observations. The estimated parameters are as follows:

- $\hat{\beta}_0 = 2.31$
- $\hat{\beta}_1 = 1.15$
- $s_{\hat{\beta}_0} = 0.057$
- $s_{\hat{\beta}_1} = 0.043$

The following hypothesis test is performed:

- $H_0 : \beta_1 = 1$
- $H_1 : \beta_1 \neq 1$

Calculate the minimum significance level at which the null hypothesis would be rejected.

✖ Incorrect Answer

12% A Less than 0.01

13% ✖ At least 0.01, but less than 0.02

64% ✔ At least 0.02, but less than 0.05

7% D At least 0.05, but less than 0.10

3% E At least 0.10

15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33

You are given:

- An actuary wants to study the effect of birth season on the height of individuals.
- Her data set consists of the following 12 individuals:

Individual	Season of Birth	Height		Individual	Season of Birth	Height
1	Spring	155.4		7	Fall	172.7
2	Spring	142.6		8	Fall	165.2
3	Spring	150.6		9	Fall	180.3
4	Summer	155.3		10	Winter	175.6
5	Summer	161.6		11	Winter	172.5
6	Summer	157.0		12	Winter	172.0

- She fits an ordinary least squares regression model using Summer as the reference category.
- Below is the partial R output for her regression model:

Coefficients	Estimate
(Intercept)	b_0
SeasonSpring	-8.433
SeasonFall	b_2
SeasonWinter	b_3

Determine which of the following statements about the coefficient estimates can be concluded correctly from the given information.

6% A $b_0 = 163.4$

5% B $b_2 = 9.333$

63% ✓ $b_3 = 15.4$

12% D All (A), (B), and (C) can be concluded correctly from the given information.

14% E There is not enough information to infer (A), (B), or (C).

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

A botanist wants to investigate the effects of pesticides on the number of strawberries grown. He designed an experiment with three strawberry patches (each with ten plants) where one patch serves as a control, and the other two treated with pesticides A and B, respectively.

After six months, the number of strawberries are recorded. Below are the observations:

$$\sum_{i=1}^{3}\sum_{j=1}^{10}(y_{ij}-\bar{y}_i)^2 = 534 \qquad \sum_{i=1}^{3}\sum_{j=1}^{10}(y_{ij}-\bar{y})^2 = 742$$

where:

- y_{ij} is the number of strawberries grown by plant j in patch i.
- $\bar{y}_i$ is the mean number of strawberries grown in patch i.
- $\bar{y}$ is the overall mean number of strawberries grown.

The botanist wants to test if pesticides have any impact on the growth of strawberries. He formulates the following hypothesis:

- H_0 : The pesticides do not have a significant effect on the growth of strawberries.
- H_1 : The pesticides have a significant effect on the growth of strawberries.

Calculate the F statistic to test the null hypothesis.

✖ Incorrect Answer

17% ✖ Less than 4.0

5% B At least 4.0, but less than 4.5

8% C At least 4.5, but less than 5.0

61% ✔ At least 5.0, but less than 5.5

9% E At least 5.5

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

In an experiment with three blocks, three treatments, and a total of nine observations you are given the following summary of results:

	Sum of Squares	Mean Square
Between Treatments	$\frac{146}{9}$	$\frac{73}{9}$
Between Blocks	$\frac{50}{9}$	$\frac{25}{9}$
Residual	$\frac{10}{9}$	$\frac{5}{18}$
Total	$\frac{206}{9}$	

- H_0 : The treatment means are all equal.
- H_1 : The treatment means are not all equal.
- A block is defined to be a group of three observations.
- The results are Normally distributed.

Calculate the smallest p-value for which H_0 can be rejected.

70% ✔ Less than 1%

7% B At least 1%, but less than 2.5%

13% C At least 2.5%, but less than 5%

4% D At least 5%, but less than 10%

6% E At least 10%

29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

Two different data sets were used to construct the four regression models below. The following output was produced from the models:

Data Set	Model	Dependent variable	Independent variables	Total Sum of Squares	Residual Sum of Squares
A	1	Y_A	X_{A1}, X_{A2}	35,930	2,823
A	2	X_{A1}	X_{A2}	92,990	7,070
B	3	Y_B	X_{B1}, X_{B2}	27,570	13,240
B	4	X_{B1}	X_{B2}	87,020	34,650

Determine which one of the following statements best describe the data.

✖ Incorrect Answer

11% A Collinearity is present in both data sets A and B.

12% B Collinearity is present in neither data set A nor B.

69% ✔ Collinearity is present in data set A only.

6% ✖ Collinearity is present in data set B only.

3% E The degree of collinearity cannot be determined from the information given.

29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

You are given:

- A linear model, $Y = \beta_0 + \beta_1 x + \varepsilon$, is fitted to a data set.
- The data set has 250 observations.
- $\sum_{i=1}^{250} x_i = 87{,}391$
- $\sum_{i=1}^{250} x_i^2 = 36{,}268{,}085$

Calculate the leverage statistic for an observation with $x = 1{,}500$.

Incorrect Answer

15% A Less than 0.10

6% At least 0.10, but less than 0.20

73% At least 0.20, but less than 0.30

4% D At least 0.30, but less than 0.40

2% E At least 0.40

23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42

A normal regression with identity link is performed on a set of data to investigate the average credit card debt of a person. The following tables are results of the regression:

Response variable	Average Credit Card Debt
Response distribution	Normal
Link	Identity
Residual Std. Error	102.9

Parameter	$\hat{\beta}$	Standard Error
Intercept	-547.31	21.4604
Income	-7.80	0.2422
Student		
No	0.00	0.0000
Yes	417.51	17.1716
Rating	3.98	0.05458
Age	-0.62	0.30407

You are also given the following information on one of the observations:

Income	Student	Rating	Age	Average Debt	Leverage
106.25	Yes	483	82	903	0.0377

Calculate the Cook's distance for this observation.

✖ Incorrect Answer

65% ✔ Less than 0.0001

5% ✖ At least 0.0001, but less than 0.0003

8% C At least 0.0003, but less than 0.0005

5% D At least 0.0005, but less than 0.0007

17% E At least 0.0007

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

You are considering using k-fold cross-validation (CV) in order to estimate the test error of a regression model, and have two options for choice of k:

- 5-fold CV
- Leave-one-out CV (LOOCV)

Determine which one of the following statements makes the best argument for choosing LOOCV over 5-fold CV.

4% A 1-fold CV is usually sufficient for estimating the test error in regression problems.

8% B LOOCV and 5-fold CV usually produce similar estimates of test error, so the simpler model is preferable.

8% C Running each cross-validation model is computationally expensive.

This is an incorrect answer

67% ✓ Models fit on smaller subsets of the training data result in greater overestimates of the test error.

13% E Using nearly-identical training data sets results in highly-correlated test error estimates.

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

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

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

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

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

Determine which of the above statements are true.

6% A I only

8% B II only

72% ✔ III only

8% D I, II and III

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

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

You want to investigate whether a particular day of the week affects the number of hourly users of a bike sharing program in your state.

Determine which of the following distribution and link function is most appropriate for this model.

6% A Normal distribution, identity link function

2% B Normal distribution, log link function

9% C Binomial distribution, logit link function

76% ✓ Poisson distribution, log link function

7% E Poisson distribution, identity link function

32
33
34
35
36
37
38
39
40
41
42
43
44
45

Andy wants to study the impact of the number of miles driven on the frequency of personal auto insurance claims.

Determine which of the following distribution and link function is most appropriate for his model.

10% A Normal distribution, identity link function

5% B Binomial distribution, logit link function

72% ✓ Poisson distribution, log link function

8% D Poisson distribution, identity link function

5% E Gamma distribution, negative inverse link function

32
33
34
35
36
37
38
39
40
41
42
43
44
45

Determine which of the following statements are true.

I. The deviance is useful for testing the significance of explanatory variables in nested models.
II. The deviance for normal distributions is proportional to the residual sum of squares.
III. The deviance is defined as a measure of distance between saturated and fitted model.

6% A I only

3% B II only

13% C III only

14% D All but III

64% ✓ All

32
33
34
35
36
37
38
39
40
41
42
43
44
45

A GLM was used to estimate the expected losses per customer across gender and territory. The following information is provided:

- The link function selected is log.
- Q is the base level for Territory.
- Male is the base level for Gender.
- Interaction terms are included in the model.

The GLM produced the following predicted values for expected loss per customer:

	Q	R
Male	148	545
Female	446	4,024

Calculate the estimated beta for the interaction of Territory R and Female.

5% A Less than 0.85

66% At least 0.85, but less than 0.95

16% C At least 0.95, but less than 1.05

7% D At least 1.05, but less than 1.15

6% E At least 1.15

32
33
34
35
36
37
38
39
40
41
42
43
44
45

Three generalized linear models are built to predict the sepal length of irises. Details for these three models are provided below:

Model I	
Distribution	Normal
Link	Identity

Parameter	Estimate	p-value
(Intercept)	1.856	9.85E-12
Sepal.Width	0.6508	< 2e-16
Petal.Length	0.7091	< 2e-16
Petal.Width	-0.5565	2.41E-05

Model Statistics	
Deviance	14.445
AIC	84.643

Model II	
Distribution	Inverse Gaussian
Link	Inverse-square

Parameter	Estimate	p-value
(Intercept)	0.0639	< 2e-16
Sepal.Width	-0.0047	6.80E-11
Petal.Length	-0.006	< 2e-16
Petal.Width	0.0033	4.33E-03

Model Statistics	
Deviance	0.080
AIC	95.898

Model III	
Distribution	Gamma
Link	Inverse

Parameter	Estimate	p-value
(Intercept)	0.2795	< 2e-16
Sepal.Width	-0.0161	5.96E-14
Petal.Length	-0.0189	< 2e-16
Petal.Width	0.0123	4.01E-04

Model Statistics	
Deviance	0.426
AIC	82.629

The Akaike Information Criterion is used to select the best out of these three models.

Calculate the predicted sepal length for an iris with a sepal width of 4.4, a petal length of 6.9, and a petal width of 2.5.

9.17 (round to the nearest 0.01)

Correct Answer: 9.17

32
33
34
35
36
37
38
39
40
41
42
43
44
45

Determine which of the following link functions are suitable for a generalized linear model used to model the likelihood of a fraudulent claim. (select all that apply)

62% ✓ Complementary log-log link function

14% B Identity link function

13% C Inverse link function

82% ✓ Logit link function

6% E Power link function

32
33
34
35
36
37
38
39
40
41
42
43
44
45

A Poisson regression with identity link is performed to predict the annual number of claims. The table below shows the estimated parameters:

Response Variable	Annual Number of Claims
Response Distribution	Poisson
Link	Identity

Parameter	**df**	$\hat{\beta}$	p**-value**
Intercept	1	0.712	<0.0001
Age	1	0.003	<0.0001
Gender: Male	1	-1.122	<0.0001
Gender: Female	0	0.000	
Number of Legal Drivers in Household	1	0.465	<0.0001
Zone: A	0	0.000	
Zone: B	1	0.468	<0.0001
Zone: C	1	1.182	<0.0001

Calculate the variance of the annual number of claims for a male age 35 from zone A with 3 legal drivers in his household.

5% A Less than 1.0

75% ✓ At least 1.0, but less than 2.0

17% C At least 2.0, but less than 3.0

2% D At least 3.0, but less than 4.0

2% E At least 4.0

32
33
34
35
36
37
38
39
40
41
42
43
44
45

You have fit a generalized linear model using a Poisson distribution. One observation of the response variable is 59. The corresponding GLM fitted value is 71. The corresponding leverage of this point is 22%.

Calculate the standardized deviance residual corresponding to this observation.

✖ Incorrect Answer

13% A Less than -1.8

7% ✖ At least -1.8, but less than -1.7

58% ✔ At least -1.7, but less than -1.6

6% D At least -1.6, but less than -1.5

16% E At least -1.5

32
33
34
35
36
37
38
39
40
41
42
43
44
45

Suppose we have a linear regression model

$$Y = \beta_0 + \beta_1 \cdot b_1(X) + \beta_2 \cdot b_2(X) + \varepsilon$$

with basis functions $b_1(X) = X$ and $b_2(X) = X^2$. This model can be rewritten as

$$Y = \alpha_0 + \alpha_1 \cdot b_1^*(X) + \alpha_2 \cdot b_2^*(X) + \varepsilon$$

where $b_1^*(X) = X - 1$ and $b_2^*(X) = (X-1)^2$.

Which of the following is the correct expression for α_0 in terms of the β parameters?

✖ Incorrect Answer

3% A $\beta_0 + 2\beta_1$

67% ✔ $\beta_0 + \beta_1 + \beta_2$

17% C $\beta_0 + 2\beta_1 + \beta_2$

5% D $\beta_0 + 2\beta_1 + 2\beta_2$

8% ✖ The answer is not given by (A), (B), (C) or (D).

32
33
34
35
36
37
38
39
40
41
42
43
44
45

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.

2% A Add a linear constraint in the regions before and after the first knot

1% B Increase the number of orders in the regression equation

5% C Increase the number of knots in the model

24% D Reduce the number of knots in the model

67% ✓ Increase the span, s, of the model

23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43

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	24.7	21.9	26.4	25.5	28.2	29.1	26.7	27.1
x_i	15.2	15.4	15.9	16.4	17.5	18.8	19.0	19.4	22.8	26.5

When fitting the linear local regression at $X = 10$, you are using the weighting function

$$K_i = \frac{2 - |10 - x_i|}{2}$$

This linear local regression at $X = 10$ has produced the estimate $\hat{\beta}_1 = 2.489$.

Calculate the estimated value of $\hat{\beta}_0$.

✖ Incorrect Answer

8% A Less than -13

9% ✖ At least -13, but less than -11

73% ✔ At least -11, but less than -9

6% D At least -9, but less than -7

4% E At least -7