CA – Exam 2

Exam Results

Nov 2nd, 2024 at 7:15 PM

Back to Practice

Score 31/45	Difficulty	Earned Level	Total Time
69%	4.3	4.1 +0.4	4:00:18 of 4:00:00

Section Review

1 Probability Models	9/15	60%	1:38:20	33%
2 Statistics	8/8	100%	44:12	18%
3 Extended Linear Models	14/22	64%	1:37:46	49%

0/1 | 62% | 1.1 | 5.1 | 4:01 | 4:58

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

You are given the following information about random variable X:

- The hazard rate function is:

$$r(x) = \begin{cases} \frac{k^2}{3x}, & \text{for } x \geq 2 \\ 0, & \text{otherwise} \end{cases}$$

- The value of the cumulative distribution function at $x = 5$ is:

$$F(5) = 0.936$$

Calculate the absolute value of k.

Incorrect Answer

8% A Less than 1.5

17% B At least 1.5, but less than 2.5

62% C At least 2.5, but less than 3.5

5% D At least 3.5, but less than 4.5

7% E At least 4.5

1/1 | 61% | 1.2 | 5.4 | 7:50 | 4:39

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

Losses in 2008 follow an exponential distribution with mean 100. Each loss is subject to an ordinary deductible of 10.

In 2009, the deductible is removed and replaced with a coinsurance of factor of α.

Inflation is 10% per year, which affects all losses. The expected payment per loss is the same in both years.

Calculate α.

1% A Less than 0.80

61% B At least 0.80, but less than 0.85

6% C At least 0.85, but less than 0.90

25% D At least 0.90, but less than 0.95

6% E At least 0.95

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

Losses follow an exponential distribution with parameter θ.

For a deductible of 100, the expected payment per loss is 2,000.

Determine the expected payment per loss for a deductible of 500.

Incorrect Answer

8% A θ

11% $\theta\left(1-e^{-500/\theta}\right)$

60% $2{,}000e^{-400/\theta}$

5% D $2{,}000e^{-5/\theta}$

15% E $2{,}000 \cdot \dfrac{1-e^{-500/\theta}}{1-e^{-100/\theta}}$

0/1 64% 1.3 5.9 4:37 7:52

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

For a Pareto distribution with $\alpha = 3$ and θ, the absolute difference between the 90th percentile and the 99th percentile is 497.431.

Let CTE_δ be the conditional tail expectation with $1 - \delta$ tolerance probability.

Determine the absolute difference between the $\mathrm{CTE}_{0.90}$ and the $\mathrm{CTE}_{0.99}$.

Incorrect Answer

10% A Less than 600

12% At least 600, but less than 700

64% At least 700, but less than 800

8% D At least 800, but less than 900

5% E At least 900

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

You are given the following information about a senior-year college student from an actuarial science program:

- Job offers arrive according to a Poisson process with a rate of three per month.
- A job offer is acceptable if the salary offered is at least 50,000 per year.
- The salaries offered are mutually independent and follow a lognormal distribution with $\mu = 10.5$ and $\sigma = 0.2$.

Calculate the probability that it will take this senior-year college student more than four months to receive an acceptable job offer.

Incorrect Answer

13% A Less than 0.505

16% At least 0.505, but less than 0.515

61% At least 0.515, but less than 0.525

4% D At least 0.525, but less than 0.535

6% E At least 0.535

1/1 60% 1.4 4.9 14:16 3:57

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

A customer calls a customer service center with two servers. Both servers are busy at the time of the call, therefore the customer is placed in a queue that is currently empty. The next available server will handle this customer's call.

You are given:

- Service times are exponentially distributed and independent.
- Server 1 is more experienced, so she handles calls in 5 minutes on average.
- Server 2 is less experienced, so she handles calls in 7 minutes on average.

Calculate the expected total waiting plus service time, in minutes, until this customer is finished.

10% A Less than 7

5% B At least 7, but less than 8

60% At least 8, but less than 9

8% D At least 9, but less than 10

18% E At least 10

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

You are given the following information on a health insurance policy:

- Claims occur according to a Poisson process with mean $\lambda = 10$ per week.
- Claim amounts follow the Pareto distribution with probability density function
- $f(x) = \dfrac{4 \cdot 1{,}000^4}{(x + 1{,}000)^5} \quad , 0 < x$

Calculate the probability that the total claim amount in 52 weeks exceeds 200,000 using the Normal approximation.

13% A Less than 0.02

76% B At least 0.02, but less than 0.03

5% C At least 0.03, but less than 0.04

3% D At least 0.04, but less than 0.05

3% E At least 0.05

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

You are given the following information for a policy covering two types of claims:

- Total number of claims is given by a Poisson process with claims intensity $\lambda(t) = 10t, t > 0$.
- At any time the probability of a claim being from Claim Type A is 0.6 and from Claim type B is 0.4.
- Frequency and severity of claims are independent.
- Claim severities follow the distributions given in the tables below.

Claim Type A

Claim Amount	Probability
< 500	0.3
At least 500, but less than 1,000	0.5
At least 1,000	0.2

Claim Type B

Claim Amount	Probability
$< 1{,}000$	0.1
At least 1,000, but less than 2,000	0.6
At least 2,000	0.3

Calculate the probability that by time 0.5 there will be fewer than two claims with severity at least equal to 1,000.

18% A Less than 0.55

9% B At least 0.55, but less than 0.65

11% C At least 0.65, but less than 0.75

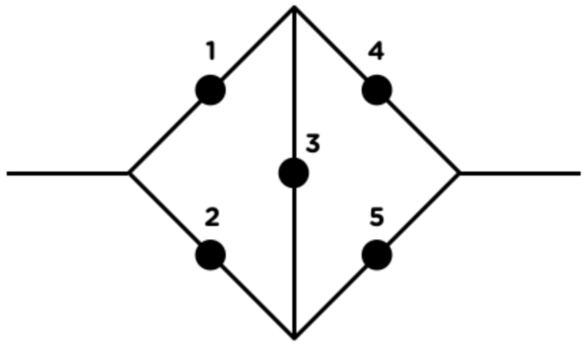
6% D At least 0.75, but less than 0.85

56% At least 0.85

1/1 55% 1.5 5.4 2:25 1:31

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

Consider the following bridge system:



Determine which of the following is a minimal cut set of the system. (select all that apply)

11% A {1, 4}

25% B {1, 2, 3}

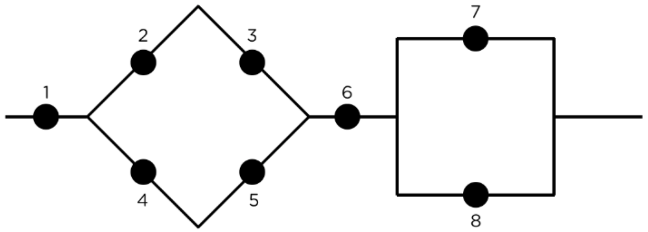
81% ✔ {1, 3, 5}

81% ✔ {2, 3, 4}

6% E {2, 3, 5}

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

You are given a system whose structure is represented in the diagram below:



The structure has identical components, with probabilities of working all equal to 0.6.

u is the upper bound reliability of the system by using the first two inclusion-exclusion bounds, defining the events in terms of minimal path sets.

Calculate u.

8% A Less than 0.150

17% B At least 0.150, but less than 0.250

68% At least 0.250, but less than 0.350

5% D At least 0.350, but less than 0.450

2% E At least 0.450

0/1 62% 1.5 6.0 8:45 4:12

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

You are given a system which consists of the following minimal cut sets:

$$\{1\}, \{2, 3\}, \{4\}, \{5, 6\}$$

The system is comprised of independent and identically distributed components, each with reliability 0.9.

Calculate the lower bound of the reliability of the system by using the first two inclusion-exclusion bounds from the method of inclusion and exclusion.

Incorrect Answer

6% A Less than 0.75

6% B At least 0.75, but less than 0.77

62% C At least 0.77, but less than 0.79

22% D At least 0.79, but less than 0.81

4% E At least 0.81

0/1 55% 1.6 5.8 7:35 6:36

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

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

Calculate the long run probability of being in State 0.

Incorrect Answer

55% Less than 0.1

9% B At least 0.1, but less than 0.2

14% C At least 0.2, but less than 0.3

6% D At least 0.3, but less than 0.4

15% At least 0.4

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

You are given the following information about when a fair coin is flipped:

- If the outcome is Heads, 1 chip is won.
- If the outcome is Tails, 1 chip is lost.
- A gambler starts with 20 chips and will stop playing when he either has lost all his chips or he reaches 50 chips.
- Of the first 10 flips, 7 are Heads and 3 are Tails.

Calculate the probability that the gambler will lose all of his chips, given the results of the first 10 flips.

23% A Less than 0.5

63% At least 0.5, but less than 0.6

7% C At least 0.6, but less than 0.7

2% D At least 0.7, but less than 0.8

4% E At least 0.8

1/1 | 71% | 1.7 | 2.9 | 11:32 | 7:50

You are given the following information:

- A 3-year term insurance policy on a life age (80) provides for a death benefit, payable at the end of the year of death.
- The death benefit is $1{,}000t$, $t \in \{1, 2, 3\}$, where t is the year of death.
- This policy is purchased by a single premium, P, at time 0.
- If (80) lives to age 83, the single premium is returned without interest.
- Mortality rates follow the Illustrative Life Table.
- $i = 0.10$

Calculate P using the equivalence principle.

12% A Less than 830

4% B At least 830, but less than 850

7% C At least 850, but less than 870

6% D At least 870, but less than 890

71% At least 890

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

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

- X has density function: $f(x) = 12x(1-x)^2$, for $0 < x < 1$.
- The rejection method is based on $g(x) = 1$, for $0 < x < 1$.
- The rejection procedure is as follows:
  - Step 1: Generate independent random numbers Y and U, which are both uniform on (0,1).
  - Step 2: If rejection function, $h(Y)$, is true, stop and set $X = Y$. Otherwise return to Step 1.

Determine which of the following is a form of the rejection function $h(Y)$.

74% (A) $U \leq \frac{27}{4} Y(1-Y)^2$

6% B $U \geq \frac{27}{4} Y(1-Y)$

5% C $U \leq 16Y^2(1-Y)^2$

5% D $U \geq \frac{27}{4} Y^2(1-Y)^2$

11% E The answer is not given by (A), (B), (C), or (D).

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

You are given:

- Assume losses follows a single-parameter Pareto distribution.
- $\theta = 150$
- Losses given are 200, 300, 350, 400.

Find the Maximum Likelihood Estimate of α.

8% A Less than 0.8

12% B At least 0.8, but less than 1.2

73% At least 1.2, but less than 1.6

4% D At least 1.6, but less than 2.0

2% E At least 2.0

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

Suppose that a random draw is made from a normal distribution $N\left(\mu, \sigma^2\right)$, where σ is known to be 2.

The normal distribution density is given by:

$$f\left(x; \mu, \sigma\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty$$

Calculate the Fisher information of μ.

6% A Less than 0.2

77% At least 0.2, but less than 0.4

12% C At least 0.4, but less than 0.6

3% D At least 0.6, but less than 0.8

1% E At least 0.8

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

You are given the following information:

- A random variable, X, is uniformly distributed on the interval (0, θ).
- θ is unknown.
- For a random sample of size n, an estimate of θ is $Y_n = \max\{X_1, X_2, \ldots, X_n\}$.

Determine which of the following is a consistent estimator of θ.

57% A Y_n

13% B $\frac{Y_n}{n-1}$

11% C $Y_n(n+1)$

1% D $Y_n(n+1)(n-1)$

19% E $\frac{Y_n}{n+1}$

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

For Jones's political poll the question is what is the probability, p, that a voter would vote for Jones:

- 15 voters were sampled to test:

  $H_0 : p = 0.5$ vs. $H_1 : p < 0.5$

- The test statistic Y is the number of sampled voters favoring Jones.
- RR $= \{y \leq 3\}$ is selected as the rejection region.
- Assume a binomial distribution for the probability that a voter will favor Jones.

Calculate the probability of a Type I error, α.

6% A Less than 0.010

20% B At least 0.010, but less than 0.015

69% At least 0.015, but less than 0.020

2% D At least 0.020, but less than 0.025

2% E At least 0.025

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

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from a normal distribution with unknown mean μ and known variance $\sigma^2 = 25$.

- $H_0 : \mu = 20$
- $H_1 : \mu > 20$
- Significance level $\alpha = 0.05$
- Sample mean was 21.49.

Calculate the smallest value of ***n*** that would result in rejecting H_0 using the Neyman-Pearson lemma.

74% ✔ Less than 32

8% B At least 32, but less than 33

11% C At least 33, but less than 34

4% D At least 34, but less than 35

3% E At least 35

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

You are given the following information about the number of auto claims:

	Age Group		
Gender	**Young**	**Old**	**Total**
Male	20	15	35
Female	12	22	34
Total	32	37	69

A Chi-squared hypothesis test is performed to determine if gender and age group are independent.

Calculate the p-value of this test.

7% A Less than 0.005

4% B At least 0.005, but less than 0.010

6% C At least 0.010, but less than 0.025

8% D At least 0.025, but less than 0.050

75% At least 0.050

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

The probability density of Y_k, the k^{th} order statistic for a sample of size n is:

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

Let $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$ be the order statistics of five independent observations from a uniform distribution on (0, 10).

Calculate the variance of Y_4.

6% A Less than 2

68% B At least 2, but less than 4

13% C At least 4, but less than 6

8% D At least 6. but less than 8

6% E At least 8

1/1 | 76% | 2.7 | 2.7 | 3:10 | 3:19

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

Let $Y_1 < Y_2 < \cdots < Y_{12}$ be the order statistics of a random sample from a uniform distribution [0, 1].

Calculate the probability that $0.6 < Y_{12} < 0.75$.

7% A Less than 0.010

6% B At least 0.010, but less than 0.015

7% C At least 0.015, but less than 0.020

4% D At least 0.020, but less than 0.025

76% At least 0.025

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

Simon uses a statistical learning method to estimate the number of ears of corn produced per acre of land.

He has multiple training data sets and applies the same statistical method to all of them. The results are not identical, but very similar, between the training data sets.

Which of the following best describes the statistical learning method?

70% The method has low variance.

3% B The method has high variance.

2% C The method has high training error.

18% D The method has low training error.

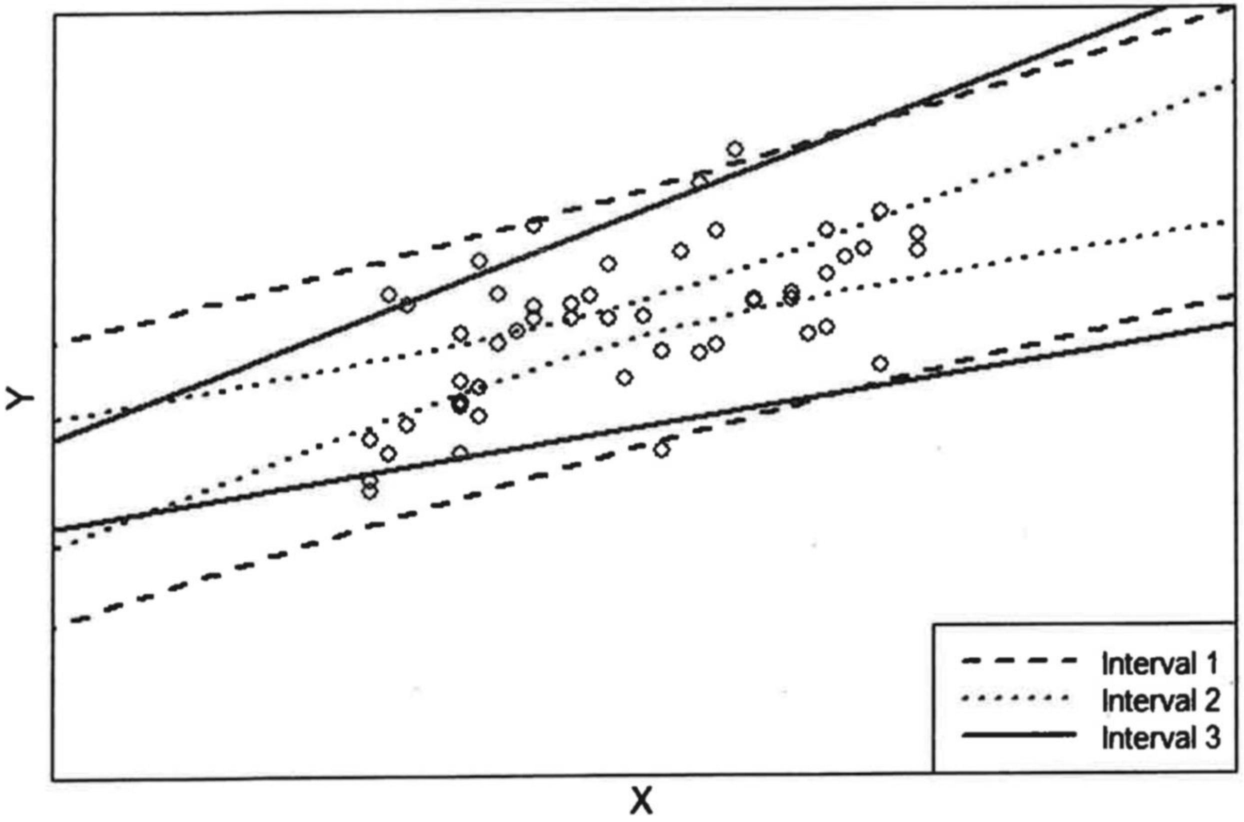
7% E The method has low bias.

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

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

$$\mathrm{E}[Y_i] = \beta_0 + \beta_1 X_i$$

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



Let "CI" be the 95% confidence interval for $\mathrm{E}[Y_i]$, and let "PI" be the 95% prediction interval for Y_i.

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

You wish to explain Y using the following multiple regression model and 32 observations:

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

A linear regression package generates the following table of summary statistics:

	Estimated Coefficient	Standard Error
Intercept	44.200	5.960
β_1	-0.295	0.118
β_2	9.110	6.860
β_3	-8.700	1.200

For the intercept and each of the β's, you decide to reject the null hypothesis which is that the estimated coefficient is zero at $\alpha = 10\%$ significance.

Which variables have coefficients significantly different from zero?

3% A Intercept

2% B Intercept, X_1

12% C Intercept, X_2

73% Intercept, X_1, X_3

10% E Intercept, X_2, X_3

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

The following R output is produced from a linear regression on a dataset with 100 observations:

Coefficients	Estimate	Standard Error
(Intercept)	33.38494	2.05294
Police	0.95909	0.82901
AccidentYes	12.65060	1.70702
Temp	-0.07916	0.02102
Departure	-0.60078	0.13615

- Police, Temp, and Departure are continuous variables.
- Accident is a binary categorical variable where No is the reference category.

Based on the output, determine which of the following predictors should be removed, if any.

9% A None of the predictors should be removed

61% Police

20% C Accident

7% D Temp

3% E Departure

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

The following two linear regression models were fit to 20 observations:

- Model 1: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
- Model 2: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$

The results of the regression are as follows:

Model Number	Error Sum of Squares	Regression Sum of Squares
1	13.47	22.75
2	10.53	25.70

The null hypothesis is $H_0 : \beta_3 = \beta_4 = 0$ with the alternative hypothesis that the two β's are not equal to zero.

Calculate the statistic used to test H_0.

8% A Less than 1.70

4% B At least 1.70, but less than 1.80

5% C At least 1.80, but less than 1.90

5% D At least 1.90, but less than 2.00

78% ✔ At least 2.00

1/1 75% 3.3 2.6 1:49 3:19

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

The following two models were fit to 18 observations:

- Model 1: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
- Model 2: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2 + \varepsilon$

The results of the regression are:

Model Number	Error Sum of Squares	Regression Sum of Squares
1	102	23
2	78	47

Calculate the value of the F-statistic used to test the hypothesis that $\beta_3 = \beta_4 = \beta_5 = 0$.

75% Less than 1.30

6% B At least 1.30, but less than 1.40

7% C At least 1.40, but less than 1.50

4% D At least 1.50, but less than 1.60

8% E At least 1.60

22
23
24
25
26
27
28
29
30
31
32
33
34
35

You are given the following information on 12 individuals:

Individual	Season of Birth	Height
1	Spring	152.5
2	Spring	157.4
3	Spring	151.5
4	Summer	164.8
5	Summer	163.6
6	Summer	163.3
7	Fall	172.7
8	Fall	170.4
9	Fall	162.2
10	Winter	171.3
11	Winter	188.2
12	Winter	182.4

You are interested in the significance of the predictor Season of Birth in predicting Height.

Calculate the F-statistic that is used to evaluate whether or not Season of Birth is significant.

Incorrect Answer

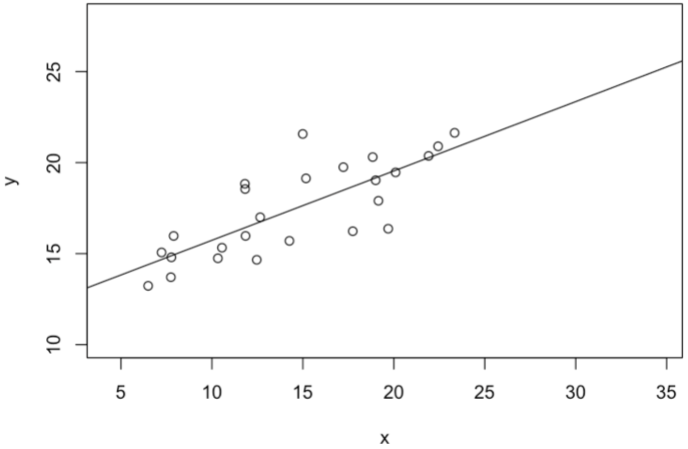
0.5 (round to the nearest 0.1)

Correct Answer: 12.9

22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

You fit a regression model to a dataset with 25 observations.

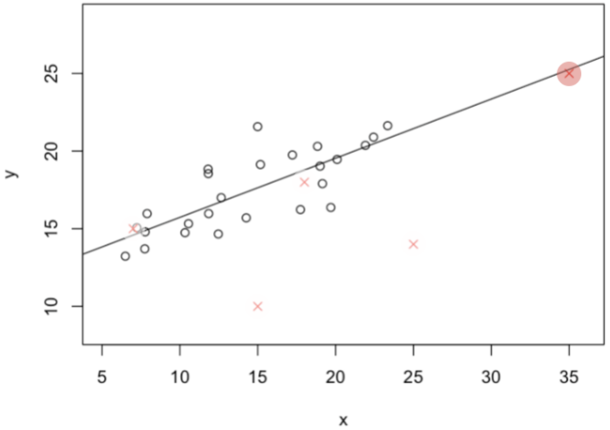
You are given the scatterplot of the dataset with the superimposed line of best fit.



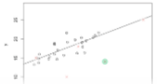
You consider adding one of five observations to the model. These five observations are indicated by red X's in the scatterplot below.

Determine which of the five observations would ***most likely*** be a high leverage point and an outlier if added to the model. (click on one of the red X's)

Incorrect Answer(s)



Correct Answer



22
23
24
25
26
27
28
29
30
31
32
33
34
35

You are given the following statements about different resampling methods:

I. Leave-one-out cross-validation (LOOCV) is a special case of k-fold cross-validation.
II. k-fold cross-validation has higher variance than LOOCV when $k < n$.
III. LOOCV tends to overestimate the test error rate in comparison to validation set approach.

Determine which of the above statements are correct.

72% I only

2% B II only

3% C III only

15% D I, II and III

9% E The correct answer isn't given by (A), (B), (C), or (D).

0/1 73% 3.7 4.2 3:55 4:06

22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

You are fitting the linear regression model

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

to a set of 10 data points. Three potential sets of coefficients and their corresponding residual sum of squares (RSS) are:

- Model I: $\hat{\beta}_0 = 22.05$, $\hat{\beta}_1 = 0.30$, $\hat{\beta}_2 = 0.67$, $\hat{\beta}_3 = 1.03$, $\text{RSS} = 91.2$
- Model II: $\hat{\beta}_0 = 27.88$, $\hat{\beta}_1 = 0$, $\hat{\beta}_2 = 0.28$, $\hat{\beta}_3 = 0.64$, $\text{RSS} = 98.8$
- Model III: $\hat{\beta}_0 = 31.94$, $\hat{\beta}_1 = 0$, $\hat{\beta}_2 = 0$, $\hat{\beta}_3 = 0.21$, $\text{RSS} = 103.3$

where:

$$\text{RSS} = \sum_{i=1}^{10} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1} - \hat{\beta}_2 x_{i,2} - \hat{\beta}_3 x_{i,3} \right)^2$$

One of the potential models is the result of lasso with tuning parameter $\lambda = 6$.

Calculate the predicted value of Y under that model when $X_1 = 4$, $X_2 = 5$, and $X_3 = 3$.

Incorrect Answer

73% Less than 30

4% B At least 30, but less than 31

13% C At least 31, but less than 32

9% At least 32, but less than 33

1% E At least 33

22
23
24
25
26
27
28
29
30
31
32
33
34
35

You are testing the addition of a new categorical variable into an existing GLM. You are given the following information:

- The change in model deviance after adding the new variable is -53.
- The change in AIC after adding the new variable is -47.
- The change in BIC after adding the new variable is -32.
- Prior to adding the new variable, the model had 15 parameters.

Calculate the number of observations in the model.

5% A Less than 1,000

64% ✔ At least 1,000, but less than 1,100

19% C At least 1,100, but less than 1,200

7% D At least 1,200, but less than 1,300

5% E At least 1,300

0/1 73% 3.8 3.5 1:39 1:25

22
23
24
25
26
27
28
29
30
31
32
33
34
35

An actuary is asked to model a non-negative response variable and requires that the model form produces an unbiased estimate.

Determine which error structure and link function combination would be the best choice for the modeling request.

Incorrect Answer

8% A Poisson and Identity

5% B Compound Poisson-Gamma and Log

6% C Normal and Identity

8% Gamma and Log

73% Poisson and Log

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

You are given the following table for model selection:

Model	Deviance Δ $(= -2\ell)$	Number of Parameters (p)	AIC
Intercept + Age	A	5	435
Intercept + Vehicle Body	392	11	414
Intercept + Age + Vehicle Value	392	X	446
Intercept + Age + Vehicle Body + Vehicle Value	B	Y	462

Calculate Y.

Incorrect Answer

7% A Less than 36

6% 36

66% 37

15% D 38

5% E At least 39

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

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

You are given the following information:

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

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

3% A Less than -1

77% At least -1, but less than 0

16% C At least 0, but less than 1

3% D At least 1, but less than 2

2% E At least 2

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

You are given the following output from a model constructed to predict the probability that a Homeowner's policy will retain into the next policy term.

Response variable		retention
Response distribution		binomial
Link		square root
Pseudo R^2		0.6521
Parameter	df	$\hat{\beta}$
Intercept	1	0.6102
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 probability that a policy with 4 years of tenure that experienced a 12% prior rate increase and has 225,000 in amount of insurance will retain into the next policy term.

Calculate the value of $\hat{\pi}$.

3% A Less than 0.60

2% B At least 0.60, but less than 0.70

76% C At least 0.70, but less than 0.80

10% D At least 0.80, but less than 0.90

10% E At least 0.90

1/1 74% 3.9 2.8 4:32 5:11

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

A bank uses a logistic model to estimate the probability of clients defaulting on a loan, and it comes up with the following parameter estimates:

i	Variable	β_i
0	Intercept	-1.6790
1	Income (in 000's)	-0.0294
2	Student [Yes]	-0.3870
3	Number of credit cards	0.7710

The following four clients applied for loans from the bank:

Client	Income	Student	# of credit cards
1	25,000	Y	1
2	10,000	Y	3
3	20,000	N	0
4	75,000	N	3

The bank will reject any loan if the probability of default is greater than 10%.

Calculate the number of clients whose loan requests are rejected.

3% A 0

5% B 1

6% C 2

74% 3

12% E 4

✖ 0/1 68% 3.10 5.5 10:28 4:53

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

A Poisson regression with log link is performed on a dataset with five observations.

$$\ln \mathrm{E}\left[Y_i\right] = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2}$$

The response and prediction of each observation are given below:

i	y_i	$\hat{\mathrm{E}}\left[Y_i\right]$
1	6.0	5.8
2	7.0	8.3
3	4.0	4.2
4	8.0	6.9
5	9.0	8.8

Determine which of the following statements is/are true.

I. The deviance residual of the fifth observation is negative.
II. The model deviance is 0.403.
III. The model Pearson chi-square statistic is 0.4.

✖ Incorrect Answer

7% A None

5% B I and II only

11% ✖ I and III only

68% ✔ II and III only

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

0/1 | 66% | 3.10 | 5.3 | 4:04 | 3:41

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

You are given, $Y_1, Y_2, \ldots, Y_n$, independent and Poisson distributed random variables with respective means μ_i for $i = 1, 2, \ldots, n$.

A Poisson GLM was fitted to the data with a log-link function expressed as:

$$\mathrm{E}\left[Y_i\right] = e^{\beta_0 + \beta_1 x_i}$$

where x_i refers to the predictor variable.

Analysis of a set of data provided the following output:

x_i	y_i	$\hat{y}_i$	$y_i \log\left(\frac{y_i}{\hat{y}_i}\right)$
0	7	6.0	1.0791
0	9	6.0	3.6492
0	2	6.0	-2.1972
1	3	6.6	-2.3654
1	10	6.6	4.1552
1	8	6.6	1.5390
1	5	6.6	-1.3882
1	7	6.6	0.4119

Calculate the observed deviance for testing the adequacy of the model.

Incorrect Answer

4% Less than 4.0

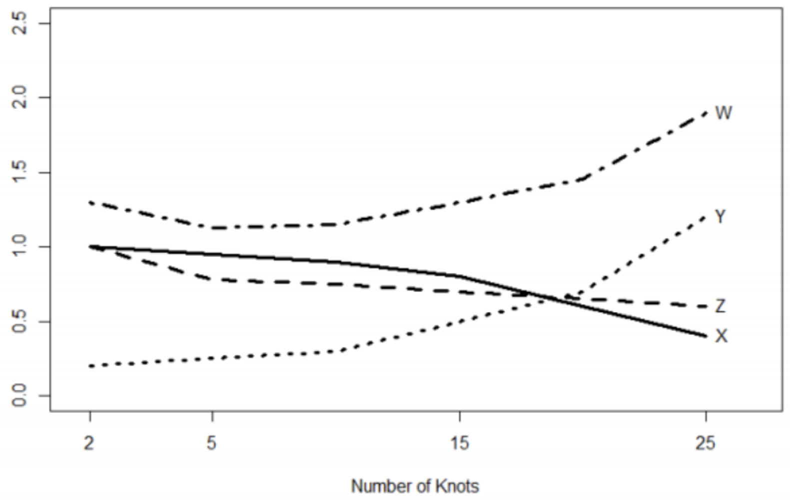
16% B At least 4.0, but less than 6.0

12% C At least 6.0, but less than 8.0

66% At least 8.0, but less than 10.0

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

You want to fit a cubic spline to a large dataset and need to determine the number of knots to use. Below is a chart of four statistics from this model valued for various numbers of knots:



Determine which set of statistics below best describes each line.

16% A W is Test MSE ; X is Variance ; Y is Squared Bias ; Z is Train MSE

4% B W is Variance ; X is Squared Bias ; Y is Test MSE ; Z is Train MSE

6% C W is Train MSE ; X is Test MSE ; Y is Variance ; Z is Squared Bias

72% D W is Test MSE ; X is Train MSE ; Y is Variance ; Z is Squared Bias

2% E W is Variance ; X is Train MSE ; Y is Test MSE ; Z is Squared Bias

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

You have fit a linear spline model to a set of data, the model having one knot. The fitted model has yielded the following predicted $\hat{y}_i$ values from the given x_i points:

$\hat{y}_i$	x_i
4.2	2
6.4	4
9.5	15
9.0	16

Determine the location of the knot.

Incorrect Answer

8% Less than 9.2

69% At least 9.2, but less than 9.5

15% C At least 9.5, but less than 9.8

4% D At least 9.8, but less than 10.1

5% E At least 10.1

1/1 | 72% | 3.11 | 3.8 | 1:00 | 1:26

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

You are fitting a model to predict the height of an adult male using shoe size as a predictor. You have a sample size of 200.

Determine which one of the following model forms will both have a discontinuity in the fitted curve and most likely overfit the data.

- 72% Piecewise cubic with two knots
- 8% B Piecewise linear with two knots
- 5% C Cubic spline with one knot
- 5% D Natural cubic spline with two knots
- 9% E Smoothing spline with 10 degrees of freedom

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

You are fitting a quadratic local regression model 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	9.1	10.2	10.9	12.1	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

Consider the following statements regarding the fitting of this model:

I. Observation 8 will be used for fitting the local regression model at the point $X = 15$.
II. A larger value of the span s would result in a smoother curve.
III. For each point fitted, at least 14 of the 20 data points will receive 0 weight.

Determine which of the above statements are true.

3% A I only

9% B II only

7% C III only

73% D I, II, and III

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