CA – Quiz – Level 1

GLMs

You are given:

- Computer lifetimes are independent and exponentially distributed with a mean of 24 months.
- Computer I has been functioning properly for 36 months.
- Computer II is a brand new and functioning computer.

Calculate the absolute difference between Computer I's failure rate and Computer II's failure rate.

81% ✓ Less than 0.01

5% B At least 0.01, but less than 0.02

9% C At least 0.02, but less than 0.03

3% D At least 0.03, but less than 0.04

2% E At least 0.04

View Solution

You are given the following information:

- Random variable X has an exponential distribution.
- $\operatorname{Var}[X] = \frac{1}{9}$

Calculate the median of the distribution.

2% A Less than 0.15

2% B At least 0.15, but less than 0.20

90% ✔ At least 0.20, but less than 0.25

1% D At least 0.25, but less than 0.30

5% E At least 0.30

View Solution

You are given:

- A sample of size n is observed from a normal distribution with variance 64.
- $H_0 : \mu = 2$
- $H_1 : \mu = 6$
- H_0 is rejected when $\bar{X} \geq 5.25$.

Calculate the smallest sample size, ***n***, needed so that the probability of a Type II error is less than 0.10.

5% A Less than 150

82% ✔ At least 150, but less than 200

7% C At least 200, but less than 250

2% D At least 250, but less than 300

5% E At least 300

View Solution

You are given the following information for a fitted GLM used to model the number of claims in a year:

Response variable	Number of claims
Response distribution	Poisson
Link	Log

Parameter	**df**	$\hat{\beta}$
Intercept	1	0.421
Zone	2	
1	0	0.000
2	1	0.484
3	1	0.292
Gender	1	
Male	0	0.000
Female	1	-0.183

Calculate the variance of the number of claims in a year for a Female in Zone 3.

1% A Less than 0.50

9% B At least 0.50, but less than 1.00

4% C At least 1.00, but less than 1.50

84% ✓ At least 1.50, but less than 2.00

You are deciding between using either regression spline or local regression to model your data. Your data has one predictor.

Determine which of the following components needs to be considered for a regression spline, but not for a local regression model.

5% A Span

81% ✓ Knots

9% C Polynomial function

3% D Weighting function

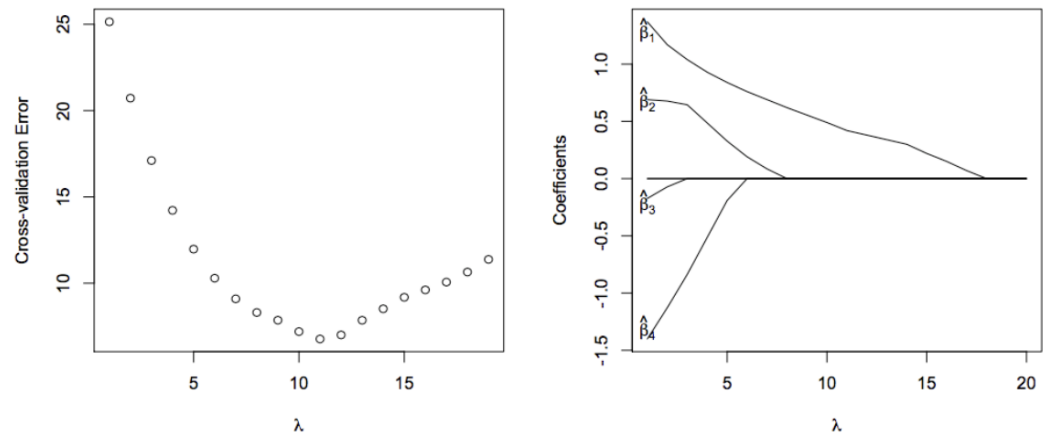
2% E The answer is not given by (A), (B), (C) or (D).

View Solution

CA – Quiz – Level 2

GLMs

You are using lasso to fit the linear regression model with four potential continuous predictors. After fitting the model, you use cross-validation to choose the best model, producing the following cross-validation and coefficient plots:



For the model chosen, the estimated intercept was $\hat{\beta}_0 = 4.6$.

Using the chosen model, estimate the predicted value of the response variable when $X_1 = 4.2, X_2 = 3.3, X_3 = 5.4, X_4 = 2.5$.

3% A Less than 6

80% At least 6, but less than 7

6% C At least 7, but less than 8

7% D At least 8, but less than 9

3% E At least 9

View Solution

You are given the following information:

- A statistician uses two models to predict the probability of success, π, of a binomial random variable.
- One of the models uses a logistic link function and the other uses a probit link function.
- There is one predictor variable, X, and an intercept in each model.
- Both models happen to produce same coefficient estimates $\hat{\beta}_0 = 0.02$ and $\hat{\beta}_1 = 0.3$.
- You are interested in the predicted probabilities of success at $x = 4$.

Calculate the absolute difference in the predicted values from the two models at $x = 4$.

4% A Less than 0.1

81% At least 0.1, but less than 0.2

5% C At least 0.2, but less than 0.3

4% D At least 0.3, but less than 0.4

5% E At least 0.4

View Solution

A Poisson regression model with log link is used to estimate the number of diabetes deaths. The parameter estimates for the model are:

Response Variable	Number of Diabetes Deaths
Response Distribution	Poisson
Link	Log

Parameter	**df**	$\hat{\beta}$	**p-value**
Intercept	1	-15.000	< 0.0001
Gender: Female	1	-1.200	< 0.0001
Gender: Male	0	0.000	
Age	1	0.150	< 0.0001
Age^2	1	0.004	< 0.0001
Age x Gender: Female	1	0.012	< 0.0001
Age x Gender: Male	0	0.000	

Calculate the expected number of deaths for a population of 100,000 females age 25.

5% A Less than 3

16% B At least 3, but less than 5

75% At least 5, but less than 7

1% D At least 7, but less than 9

2% E At least 9

View Solution

A model is created with the following form:

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

where:

$$(x - \xi)_+^3 = (x - \xi)^3 \text{ if } x > \xi, \text{ and } 0 \text{ otherwise}$$

You are given the following statements:

I. $f(x)$ is continuous at ξ.
II. $f'(x)$ is continuous at ξ.
III. $f''(x)$ is continuous at ξ.

Determine which of the above statements are true.

7% A I only

6% B II only

5% C III only

78% ✓ I, II, and III

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

View Solution

Finish Quiz

You are given:

- A set of 30 observations is fitted to a quadratic spline with one knot at x.

$$Y_i = \begin{cases} 17.725 + 10x_i + ax_i^2 + \varepsilon_i, & x_i < x \\ 310.83 - bx_i + 0.72x_i^2 + \varepsilon_i, & x_i \geq x \end{cases}$$

- At x, the fitted graph and its first derivative are continuous.
- Some observations and their fitted values are as follows:

i	x_i	$\hat{y}_i$
1	4.7	53.68
⋮	⋮	⋮
30	33.0	176.85

Calculate x.

✖ Incorrect Answer

74% ✔ Less than 16

7% ✖ At least 16, but less than 17

10% C At least 17, but less than 18

5% D At least 18, but less than 19

4% E At least 19

View Solution

CA – Quiz – Level 4

GLMs

Consider the following 2 models, which were fit to the same 30 observations using ordinary least squares:

Model 1

Response variable	Y
Response distribution	Normal
Link	Identity
SS Total	19,851
SS Error	2,781

Parameter	$\hat{\beta}$	df
Intercept	130.2	1
X_1	5.1	1
X_2	-4.2	1

Model 2

Response variable	Y
Response distribution	Normal
Link	Identity
SS Total	19,851
SS Error	2,104

Parameter	$\hat{\beta}$	df
Intercept	121.9	1
X_1	4.5	1
X_2	-5.9	1
X_3	3.5	1
X_4	-2.3	1

You test $H_0 : \beta_3 = \beta_4 = 0$ against the alternative hypothesis that at least one of $\beta_3, \beta_4 \neq 0$.

Calculate the smallest significance level at which you reject H_0.

✖ Incorrect Answer

5% ✖ Less than 0.01

5% B At least 0.01, but less than 0.02

72% ✔ At least 0.02, but less than 0.05

10% D At least 0.05, but less than 0.10

8% E At least 0.10

A normal regression with identity link and three continuous variables (plus an intercept) is performed on a set of 20 observations. You are given the following information on the fourth observation:

- The actual value and the estimated value of the fourth observation are $y = 27$ and $\hat{y} = 23.6394$, respectively.
- The standardized residual for the fourth observation is 0.7936.
- The Cook's distance for the fourth observation is 0.15403.

Calculate the residual standard error.

10% A Less than 5.3

6% B At least 5.3, but less than 5.4

8% C At least 5.4, but less than 5.5

5% D At least 5.5, but less than 5.6

71% ✓ At least 5.6

View Solution

You have $p = 10$ independent variables and would like to select a linear model to fit the data using the following two procedures:

- Best Subset Selection (BSS)
- Forward Stepwise Selection (FSS)

Let N_i be the maximum number of models fit by model selection procedure i.

Calculate the ratio $\frac{N_{\text{FSS}}}{N_{\text{BSS}}}$.

Incorrect Answer

5% ✖ Less than 0.005

11% B At least 0.005, but less than 0.010

11% C At least 0.010, but less than 0.050

68% ✔ At least 0.050, but less than 0.100

5% E At least 0.100

View Solution

Let $p = 5$ be the number of parameters that are estimated for a specific model of interest.
Let $m = 8$ be the maximum number of parameters that can be estimated for a related saturated model.

You are given the following values of the corresponding likelihood functions, when evaluated at their maximum likelihood estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_{\max}$:

$$L(\mathbf{b}) = 1.244 \times 10^{-7}; \qquad L(\mathbf{b}_{\max}) = 1.851 \times 10^{-6}$$

Calculate the value of the deviance statistic, D, and state the approximate sampling distribution of D.

6% A $D = -5.4$, and is chi-squared distributed with 3 degrees of freedom

11% B $D = 2.7$, and is chi-squared distributed with 3 degrees of freedom

5% C $D = 2.7$, and is chi-squared distributed with 5 degrees of freedom

74% ✓ $D = 5.4$, and is chi-squared distributed with 3 degrees of freedom

2% E $D = 5.4$, and is chi-squared distributed with 5 degrees of freedom

View Solution

Let $\hat{\pi}_i$ be an estimate of $\Pr(Y_i = 1)$ under a specific model of interest, where $p =$ the number of estimated parameters $= 6$ and $n =$ the number of observations $= 100$.

Let $\tilde{\pi}_i$ be an estimate of $\Pr(Y_i = 1)$ under the minimal model (in which the values of π_i are all equal).

You are given the following value of the log-likelihoods for the two models above, when evaluated at their maximum likelihood estimators:

$$l(\hat{\boldsymbol{\pi}}; \mathbf{y}) = -346.7; \qquad l(\tilde{\boldsymbol{\pi}}; \mathbf{y}) = -361.5$$

Which of the following statements about goodness-of-fit statistics, as relates to the above models, is FALSE?

6% A The value of the likelihood ratio test statistic, when comparing these two models, is 29.6.

75% ✓ The value of pseudo R^2, when comparing these two models, is 4.27%.

13% C The approximate sampling distribution of the likelihood ratio test statistic is chi-squared with 5 degrees of freedom.

3% D The value of the AIC statistic for the specific model of interest is 705.4.

3% E The value of the BIC statistic for the specific model of interest is 721.0.

View Solution