Quiz 3

Time limit. 30 minutes.

Collaboration and materials. You must complete this quiz individually. You may not use any materials (physical or electronic) besides both sides of one sheet of 8.5x11-inch paper with 1-inch margins and the equivalent of 10-point font.

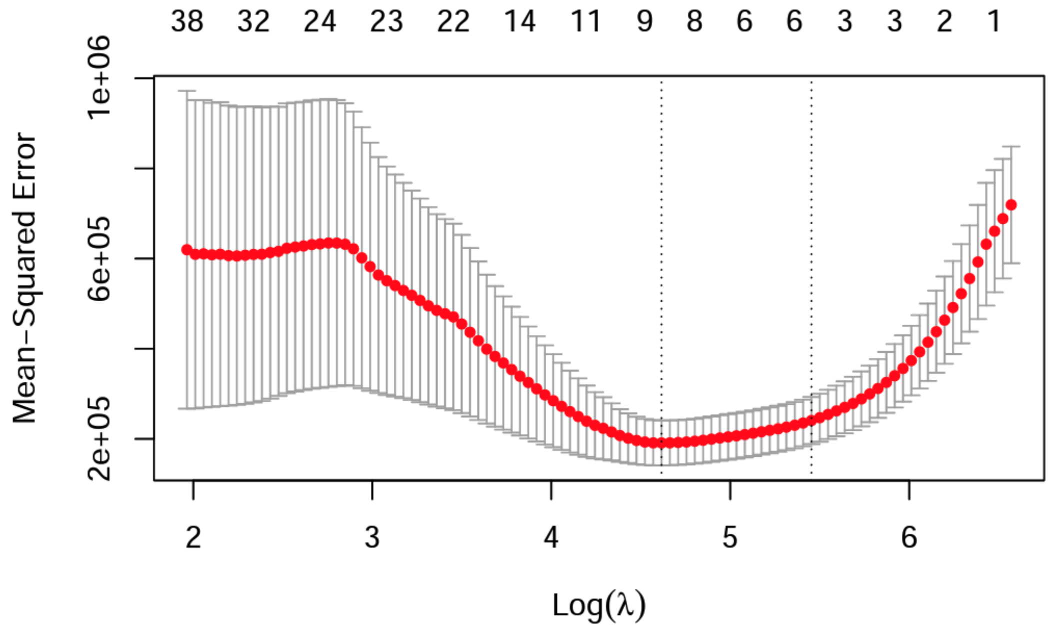
Questions. This quiz has ten multiple-choice questions. Some questions require you to select exactly one of the answer choices, while others require you to select all of the answer choices that apply. Questions of the latter kind always end with "Select all that apply."

Scoring. Each question is weighted equally. For questions requiring you to select one of the answer choices, no partial credit will be awarded. For questions requiring you to select all of the answer choices that apply, partial credit will be awarded for each correct answer selected while no points will be awarded if no correct answers are chosen or if any incorrect answers are selected.

Submission. You will receive a bubble sheet for your answers. Please print your full name as it appears on Gradescope (please no cursive), your student ID, and today's date (October 19). You may leave the "Section" box blank. **Your version is B. Please check that this matches the pre-bubbled version number at the top of the bubble sheet.** For each question, please fill in the appropriate bubbles completely using either pencil or blue/black pen. If you have filled in a bubble with pen but have changed your mind, you can cross out that bubble with an X. Note that the answer choices are presented in the order A, B, C, D, E.

1 1 point

Consider the following lasso CV plot:



Which of the following values of $\mathrm{Log}(\lambda)$ gives a CV error that is within one standard error of the minimum CV error? If there are multiple such values, choose the one that gives the sparsest model.

- 4
- 3
- 2
- 6
- 5

Typesetting math: 100%

2 1 point

Consider the L0-penalized regression with just one feature:

$$\widehat{\beta} = \underset{\beta_0,\beta_1}{\arg\min} \left\{\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 X_i))^2 + \lambda \cdot 1(\beta_1 \neq 0)\right\}.$$

Here, $1(\beta_1 \neq 0)$ is the indicator that β_1 is nonzero; it takes the value 1 if $\beta_1 \neq 0$ and 0 if $\beta_1 = 0$. For a given value of β_0, the value of $\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 X_i))^2$ can be as low as 15 when we allow β_1 to be nonzero but is 65 when we set $\beta_1 = 0$. Given this value of β_0, let $\lambda_{\max}$ be the largest value of λ for which the fitted coefficient $\widehat{\beta}_1$ will be nonzero. To which of the intervals below does $\lambda_{\max}$ belong? [Hint: Consider the values taken on by the function being minimized when $\beta_1 = 0$ and when $\beta_1 \neq 0$. The optimization will choose the value of β_1 that leads to the smaller of these two.]

- [20, 40)
- [0, 20)
- [40, 60)
- [80, 100]
- [60, 80)

3 1 point

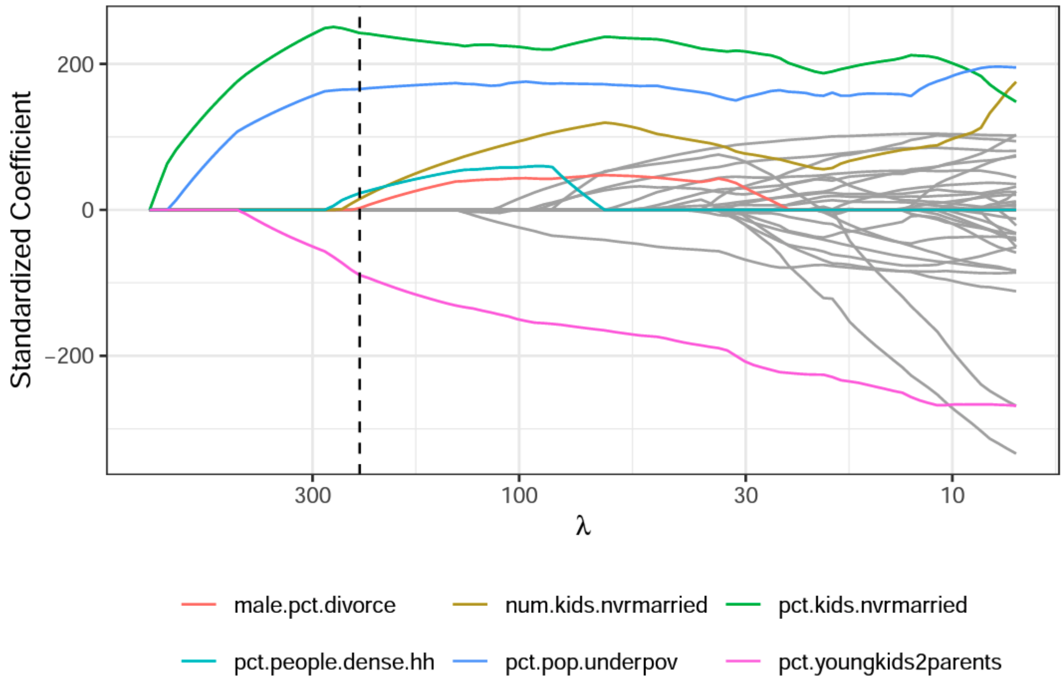
Consider a logistic regression. Which of the following quantities must lie between 0 and 1 (inclusive)? Select all that apply.

- [ ] The log of the logistic likelihood for given coefficient values
- [ ] The predicted odds that $Y = 1$ for given feature values
- [ ] The logistic likelihood for given coefficient values
- [ ] The predicted log odds that $Y = 1$ for given feature values
- [ ] The predicted probability that $Y = 1$ for given feature values

Typesetting math: 100%

4 1 point

Consider the following lasso trace plot for a regression of per capita crime rate on a number of socioeconomic variables:



According to the lasso model indicated by the vertical line, which of the features below are estimated to have an inverse relationship with per capita crime rate? Select all that apply.

- [ ] pct.people.dense.hh
- [ ] pct.kids.nvrmarried
- [ ] pct.youngkids2parents
- [ ] pct.pop.underpov
- [ ] num.kids.nvrmarried

5 1 point

Which of the regression methods below shrinks coefficients towards zero? Select all that apply.

- [ ] Ridge regression
- [ ] Logistic regression
- [ ] Linear regression
- [ ] Lasso regression
- [ ] Elastic net regression

Typesetting math: 100%

6 1 point

Suppose I run a linear regression of income (measured in dollars) on height (measured in inches). The fitted regression line is

$$\textbf{expected income} = 50,000 + 1,000 \times \textbf{height}.$$

I then decide that I prefer to measure height in feet rather than in inches, so I rerun my linear regression. I use both regression models to predict the income of someone who is six feet tall. How does the prediction change when going from the first model to the second?

- ◯ The predicted income will change, but not by a factor of 12.
- ◯ The predicted income will increase by a factor of 12.
- ◯ The predicted income will decrease by a factor of 12.
- ◯ Not enough information given.
- ◯ The predicted income will not change.

7 1 point

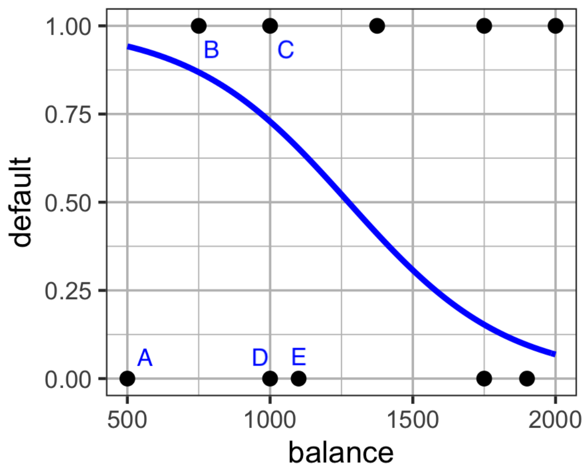
We apply ridge regression without feature standardization to a problem with four features: one is measured in inches, one in feet, one in yards, and one in miles. This regression applies the most unfairly severe penalization to the coefficients of the feature measured in which of the following units?

- ◯ Ridge regression without feature standardization treats features measured in all units fairly.
- ◯ Yards
- ◯ Feet
- ◯ Miles
- ◯ Inches

8

1 point

Shown below is a set of 10 training points as well as a candidate logistic regression fit to these points. Of the five highlighted training points, which contributes the smallest of the probabilities multiplied to calculate logistic likelihood for this candidate logistic regression fit?



- D
- B
- C
- A
- E

9

1 point

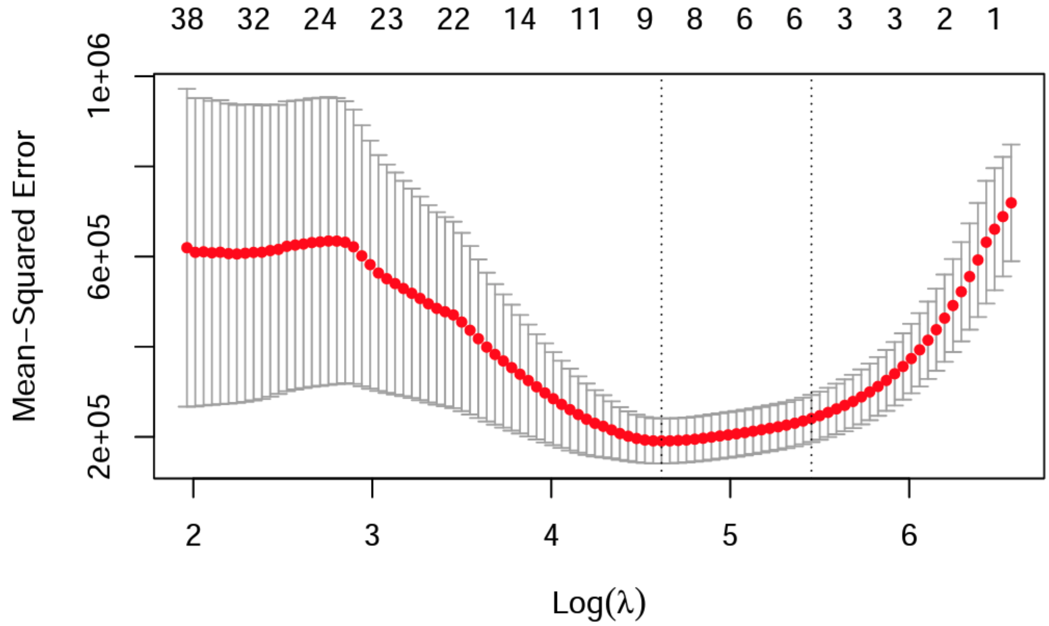
You run a logistic regression of default on student, balance, and income. Suppose the coefficient on income is -3, and income is measured in thousands of dollars. What is the correct interpretation of this coefficient?

- For fixed student status and balance, an increase in income by $1 multiplies the odds of default by e^3.
- For fixed student status and balance, an increase in income by $1 multiplies the odds of default by e^{-3}.
- For fixed student status and balance, an increase in income by $1 decreases the odds of default by 3.
- For fixed student status and balance, an increase in income by $1000 decreases the odds of default by 3.
- For fixed student status and balance, an increase in income by $1000 multiplies the odds of default by e^{-3}.

Typesetting math: 100%

10 1 point

Consider the following CV plot for a lasso regression of a response on 97 features.



Suppose an elastic net regression were applied with $\alpha = 0.5$ and the same λ as was chosen by the one-standard-error rule for the lasso regression above. Which of the choices below would be the most likely number of features with nonzero coefficients?

- 3
- 15
- 97
- 1
- 5

Typesetting math: 100%