9. (15 points) Ages

A data scientist takes a random sample of 400 people in a large city. The ages of the sampled people have an average of 35 years and an SD (standard deviation) of 20 years.

The data scientist bootstraps the sample 10,000 times, calculates the mean age of each bootstrapped sample, and finds the interval that contains the middle 95% of the 10,000 bootstrapped means. The interval goes from 33 years to 37 years.

(a) (3 pt) The interval (33 years, 37 years) is an approximate 95% confidence interval for the _____ of the people in the _____.

Fill in the blanks above by selecting from the following options.

(i) Blank 1 (make **exactly one** choice):

○ ages	○ average age	○ average	
○ sample	○ sample mean	○ city	○ city mean

(ii) Blank 2 (make **exactly one** choice):

○ ages	○ average age	○ average	
○ sample	○ sample mean	○ city	○ city mean

(b) (3 pt) The distribution of the ages of the sampled people (pick **exactly one** option):

○ is approximately normal by the Central Limit Theorem.
○ is approximately normal, but not because of the Central Limit Theorem.
○ is not normal, not even approximately.
○ may be approximately normal, or not; we need more information to decide.

(c) (3 pt) True or false: Approximately 95% of the people in the sample are between 33 and 37 years old.

○ True ○ False

(d) (3 pt) True or false: Approximately 95% of the people in the city are between 33 and 37 years old.

○ True ○ False

(e) (3 pt) The city is in a country where the average age is 35.5 years. If possible, perform a statistical test of whether or not the average age in the city is 35.5 years, using 1% as the cutoff for the p-value. State your conclusion by picking **exactly one** of the options below.

○ Since the p-value cutoff and the confidence level of the interval are inconsistent, we cannot perform this test.

○ The test concludes that the data are consistent with the hypothesis that the average age in the city is 35.5 years.

○ The test concludes that the data are not consistent with the hypothesis that the average age in the city is 35.5 years.

(f) (2.0 pt) Suppose that Gilfoyle randomly samples 100 computers from the warehouse without replacement. He observes a sample average of 570 seconds for Dogecoin mining time and he also knows that the population SD is 30 seconds.

Which of the following statements is **guaranteed** to be true?

Select all that apply.

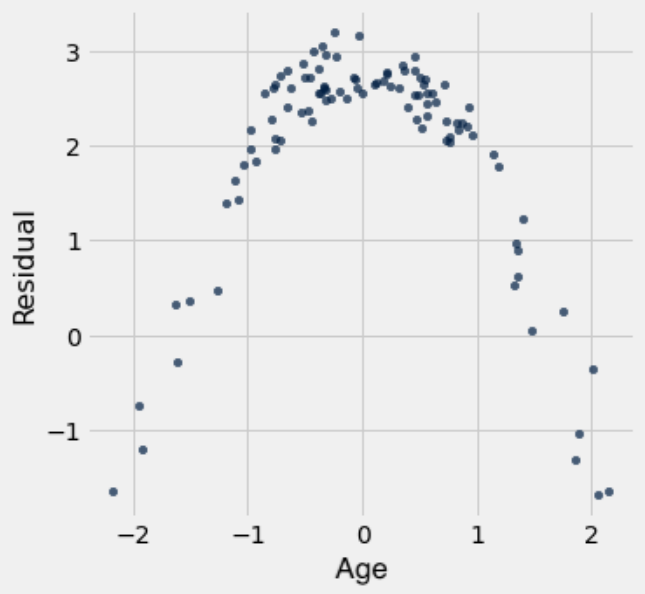
- ☐ At least 8/9ths of the computers in the population will have a Dogecoin mining time that is between 90 seconds below and 90 seconds above the population mean.
- ☐ At least 75% percent of the computers in the population will have a Dogecoin mining time that is between 510 seconds and 630 seconds.
- ☐ At least 68% percent of the computers in Gilfoyle's sample have a Dogecoin mining time that is between 540 seconds and 600 seconds.
- ☐ At least 68% percent of the computers in the population will have a Dogecoin mining time that is between 3 seconds below and 3 seconds above the population mean.
- ☐ At least 75% of the computers in the population will have a Dogecoin mining time that is between 60 seconds below and 60 seconds above the population mean.

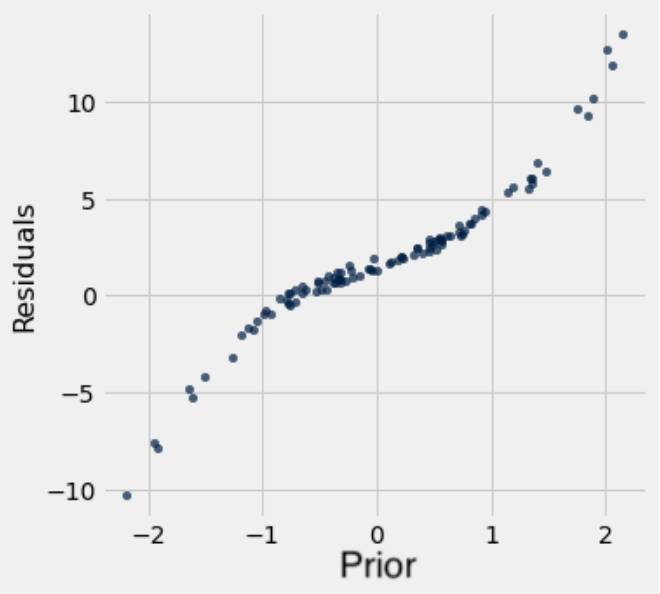
(g) (4.0 points)

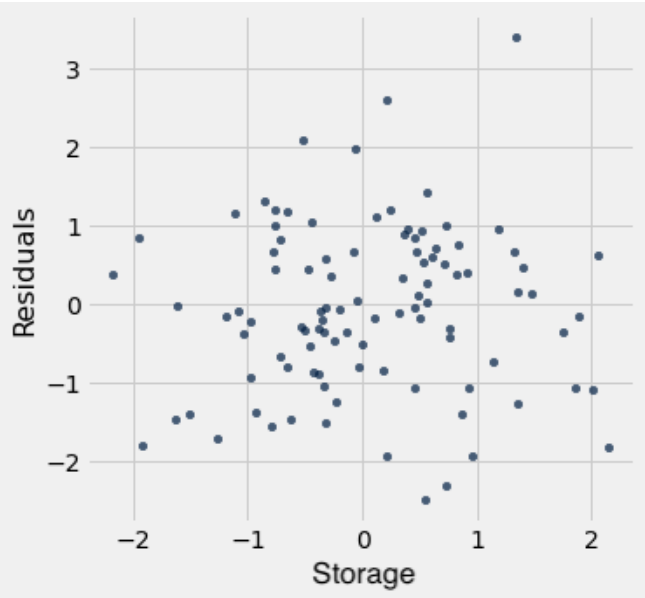
Suppose Gilfoyle tries to predict the Dogecoin mining time of his computers from each of the following four different variables:

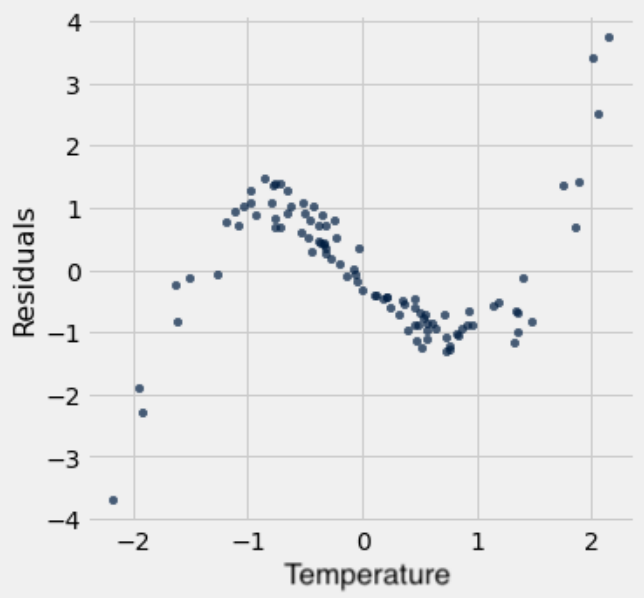
- *Age*: (int) the age of the computer, in days
- *Prior*: (float) the prior Dogecoin mining time of a computer on its most recent mine
- *Storage*: (int) the number of megabytes of storage space left on the computer
- *Temperature*: (float) the computer's temperature, in Fahrenheit

To assess his predictions from each variable, he creates the following plots:









i. **(2.0 pt)** Which of the plots above are impossible residual plots?

Select all that apply.

- ☐ *Storage.*
- ☐ *Age.*
- ☐ *Prior.*
- ☐ *Temperature.*
- ☐ None of the above.

ii. **(2.0 pt)** Which of the plots above indicate that linear regression is a good fit?

Select all that apply.

- ☐ *Temperature.*
- ☐ *Age.*
- ☐ *Prior.*
- ☐ *Storage.*
- ☐ None of the above.