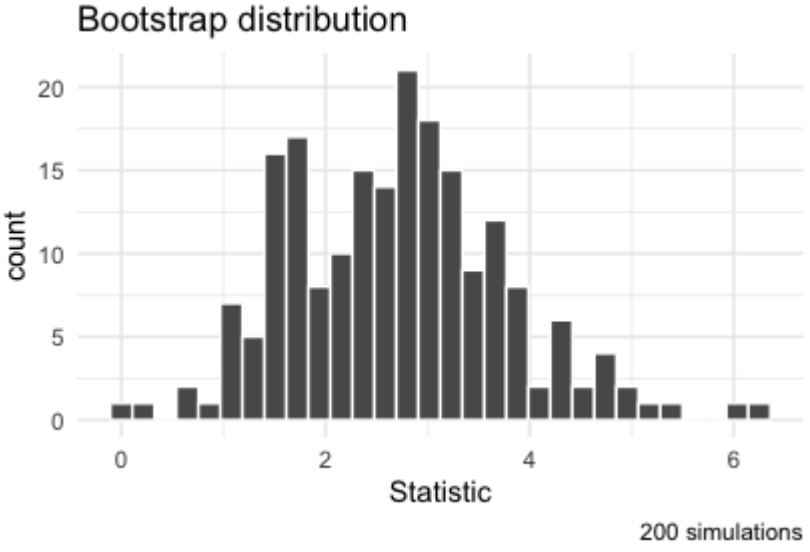
1. In a double-blind experiment a sample of male college students were asked to tap their fingers at a rapid rate. The sample was then divided at random into two groups of 10 students each. Each student drank the equivalent of about two cups of coffee, which included about 200 mg of caffeine for the students in one group but was decaffeinated coffee for the second group. After a two hour period, each student was tested to measure finger tapping rate (taps per minute). The average number of taps in the caffeine group was 246.53 and in the no caffeine group was 243.85, and both distributions were reasonably symmetric.

   (a) The goal of the experiment was to determine whether caffeine produces an increase in the average tap rate. Which of the following method(s) may be used to answer this research question? Circle **all** that apply.
   - Test for a single proportion
   - Test for a difference in proportions
   - z-test for a single mean
   - t-test for a single mean
   - t-test for a mean difference
   - t-test for a difference in means
   - Simple linear regression

   (b) We would like to calculate a 95% confidence interval for the average difference in the number of taps in the caffeine and no caffeine groups via bootstrapping. The bootstrap distribution below is created using 200 simulations. Using this distribution, estimate the 95% confidence interval, clearly state the bounds of the interval as well as marking them on the plot, and interpret your interval in context of the data.



   (c) For the test of whether caffeine produces an increase in the average tap rate, the p-value is 0.0212. Based on all of the information you have so far, which of the following intervals are plausible at a 98% confidence level for the average difference in the number of taps in the caffeine and no caffeine groups. Circle all that apply.

- $(-0.265, 4.384)$
- $(0.225, 5.269)$
- $(1.583, 6.249)$

(d) Describe in words how you would obtain the p-value in (c) using simulation-based methods.

2. Carbon monoxide (CO) for a certain kind of car vary with mean 2.9 gm/mi and standard deviation 0.6 gm/mi. A company has 100 of these cars, acquired from various (i.e., random) sources.

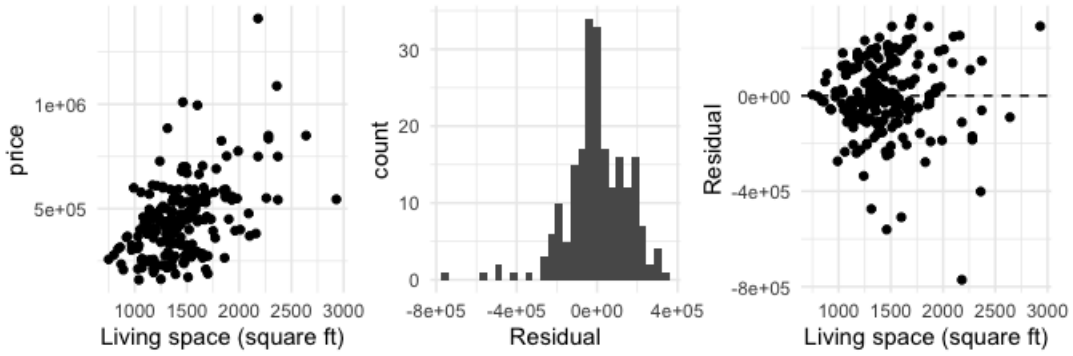
   (a) What is the probability that a randomly selected car from the fleet has CO emissions in excess of 3.1 gm/mi?

   (b) What is the probability that the average CO emissions for all 100 cars is in excess of 3.0gm/mi?

   (c) There is only a 1% chance that the company's car mean CO level is greater than what value?

3. We have data on house sale prices for King County, USA. The homes were sold between May 2014 and 2015, and we focus on 216 houses with 2 bedrooms and 2 bathrooms only. We will examine the relationship between housing prices and the square footage of the living space of the houses.

   (a) A linear model for predicting price using square footage of the living space (square feet) has been fit, and the following diagnostic plots have been produced based on this model. Which of the following is true based on these plots? Circle all that apply.



- The observations are not independent.
- The relationship between price and square footage of the house is strongly linear and positive.
- The linear model is appropriate for predicting price of a house using these data.
- The residuals are not nearly normally distributed.
- The residuals do not have equal variance.

(b) After checking the model diagnostics you realize that a linear model is not appropriate here. You decide to instead model the **log(price)** (i.e. the log price) of these houses. The model output for the log housing price using square footage as the predictor is as follows:

term	estimate	std.error	statistic	p.value
(Intercept)	12.1418376	0.1022114	118.791437	< 0.0001
sqft_living	0.0005472	0.0000683	8.008842	< 0.0001

(c) Assuming the linear model is appropriate, obtain a 90% confidence interval for β_0.

(d) Assuming the model is appropriate, what is the interpretation of the slope parameter?

(e) Someone tells you that because the estimated slope coefficient is so close to 0, there is no evidence of a relationship between the square footage of the house and its price. Is their reasoning correct? If not, explain why not.

(f) True or False: the predicted price of a house with 2000 square feet of living space is \$13.24.