

1. A local agency wants to build a playground in an abandoned field. But they must first determine if the soil is contaminated. To measure the level of contamination, they divided the field into 100 different same-sized areas and randomly chose 25 of them. Then they took some soil from each of the chosen areas and found no elevated levels of any toxic substances. What type of study is this?

Sample Survey

2. Researchers wanted to determine whether social class is related to smoking behavior. They conducted telephone interviews with 1,308 Massachusetts adolescents aged 12 to 17, selected by dialing at random. They found a statistically significant association between whether the adolescents smoked or not and the household income. Adolescents from households with less income were more likely to smoke, and this was true across all ages, for both sexes, for all races, and for all amounts of disposable income the adolescent had.

**Source:** Elpidoforos S. Soteriades and Joseph R. DiFranza. "Parent's Socioeconomic Status, Adolescents' Disposable Income, and Adolescents' Smoking Status in Massachusetts," *Journal of Public Health*, Vol. 93, July 2003, pp. 1155–1160, [www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1447926](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1447926))

- a. What type of study is this? Sample survey
- b. Can you conclude from this study that smoking is caused by an adolescent's social class? Can you think of a lurking variable that might be responsible for both?

No. A lurking variable could be a parent smokes.

- c. Can you generalize the results of this study to some larger population? Explain your thinking.

Yes, the adolescents were selected at random.

- d. Describe exactly what you can conclude from this study.

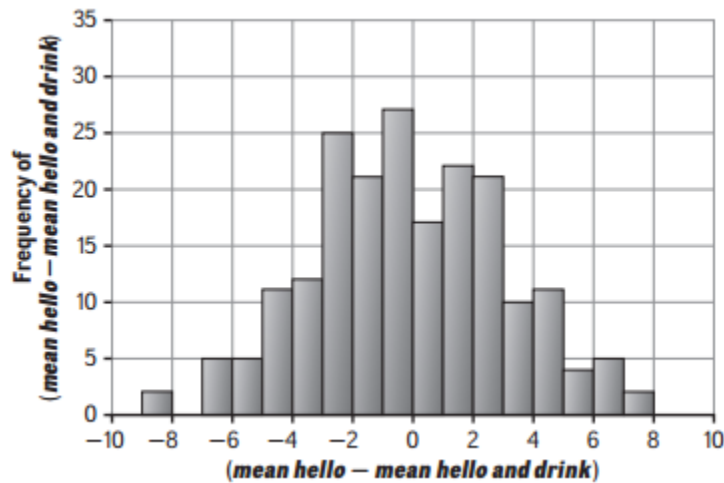
You can conclude that if you asked all adolescents aged 12 to 17 in Massachusetts, those from households with less income would be more likely to smoke than adolescents from households with more income.

3. Andrew manages a comic book store and is trying to increase sales. He decides to run an experiment for the next 12 Mondays. He randomly picks 6 of the Mondays on which he will say hello to each person as they enter the store. On the other 6 Mondays, he will say hello to each person and offer them something to drink. For each day, he will determine the percentage of people who make a purchase. The percentage of people who made purchases each day and the treatment received are given below.

Hello Only	Hello and Drink Offer
19%	14%
20%	10%
25%	26%
12%	18%
18%	27%
23%	12%

- a. What are the treatments for this experiment?  
Saying hello only or saying hello and being offered a drink
- b. What are the responses for this experiment?  
Made a purchase or did not make a purchase
- c. Is this a well-designed experiment? Explain.  
Yes. There is a control group of saying hello only. There is a sufficient number of Mondays. Since there are only 52 Mondays in a year, 12 out of 52 Mondays is large enough. There was random assignment of when he said hello.
- d. Find the value of *mean hello-mean hello and drink*.  
 $19.5 - 17.833 = 1.67$
- e. Describe how to conduct one run of a randomization test to decide whether the different treatments cause a difference in the percentage of people who make a purchase.  
Assume it makes no difference in whether Andrew says hello only or says hello and offers a drink. Write the percentages of customers who made a purchase on 12 identical slips of paper. Then randomly separate them into two groups- one for saying hello only and one for saying hello and offering a drink. Find the mean of each group. Find the difference in the means.

- f. The histogram below provides the results of 200 random assignments. The value shown in the histogram is *mean hello - mean hello and drink*.



Use the histogram to estimate the probability that if the treatments made no difference, you could get a difference just by random assignment that is at least as extreme as what occurred in the real experiment.

To estimate the probability, you need to determine how many of the runs resulted in a difference greater than 1.67 or less than -1.67. Assuming that the treatments make no difference,  $\frac{53 + 60}{200} = .565$  or 56.5%.

- g. Is the difference in the percentage of people making purchases statistically significant?

No, it is not statistically significant since the p value is greater than 5%. Therefore, saying hello only or saying hello and offering a drink will not cause different percentages of people to make a purchase.