STT 5811 HW02: Categorical Data

PUT YOUR NAME HERE

Monday, January 26, 2026 @ 04:51 PM

Contents

Instructions

Do all even-numbered exercises in **IMS2E Section 4.8**. Use complete sentences and explain your answers for those questions where you are asked to do so.

Section 4.8 Exercises

Exercise 2

a. The percent that identify themselves as conservatives is $\frac{372}{910} = 0.419 = 41.9\%$.

b. The percent that are in favor of the citizenship option is $\frac{278}{910} = 0.305 = 30.5\%$.

c. The percent that identify as conservatives *and* are in favor is $\frac{57}{910} = 0.063 = 6.3\%$.

d. The percent of approval for the citizenship option within each group is: conservatives $= \frac{57}{372} = 0.153 = 15.3\%$; moderates $= \frac{120}{363} = 0.331 = 33.1\%$; and liberals $= \frac{101}{175} = 0.577 = 57.7\%$

e. Yes, political idealogy and views on immigration seem to be associated. The percentages of conservatives, moderates, and liberals who are in favor of immigrants who entered the U.S. illegally being able to work and apply for citizenship are quite different. The more liberal the group, the larger the percentage.

f. Answers to this part may vary. In general, we are looking for variables that could reasonably be related to the other two variables being studied. For example, whether the person being polled was born in the U.S. or is an immigrant who later became a U.S. citizen might be an influential factor.

Exercise 4

a. The percentages of the three categories seem to be quite different for Democrats than Republicans and Independents/Other. This suggests the two variables may be associated.

b. Answers to this part may vary. Two examples of other variables that might explain the relationship are socioeconomic status and education level.

Exercise 6

a. The bottom graph, which has age groups as the primary bars, divided or proportioned by shipping choice, gives the preferences of shipping choices within each age group. It would be very difficult to discern differences in shipping options for 35 to 54-year-olds when looking at the top graph.

b. The top graph focuses first on each shipping option, then gives the breakdown of the age groups within each of the shipping categories.

c. Among people who are 55-plus, the most-preferred shipper is USPS (about 60%, per the second graph). Therefore, USPS will be the new company's biggest competitor, followed by UPS.

d. Among those who most prefer FedEx, the smallest age group is 55-plus (about 12%, per the first graph). Therefore, FedEx should market to that age group, followed by 35 to 54-year-olds.

Exercise 8

a. The three confounding variables are age, sex, and disease. A confounding variable is related to both of the variables being investigated: hospital and death. For example, a hospital could have older patients, and older patients might be more likely to die; a hospital might have more patients with a particular type of disease and that disease might have a higher death rate; a hospital might have more male patients, where men are more likely to die.

b. Answers to this part may vary. Additional confounding variables include socioeconomic levels of the region the hospital serves, or the distance of the hospital from the patient population it serves. The economic setting may influence hospital resources, and also will relate to the general health and rate of death for the patient population (low income individuals tend to have fewer resources generally, independent of the hospital). Distance affects who can access resources in a timely way, and may impact a hospital's connections and staffing levels.

c. If one hospital serves mostly older patients and has a higher death rate (at least in part) because it, age has confounded the relationship. The same is true if one hospital serves mostly sicker patients, with a correspondingly higher death rate. We do not know whether the actual cause of any higher death rate is the hospital itself (its practices and protocols), the ages of the patients, the level of illness, or some mix of those things. If we have not removed or controlled all of the potential confounding differences between hospitals, we cannot conclude that there is a causal association between the variables of interest. (However, we can use statistical modeling to explore how these different variables connect to one another and contribute to patient mortality.)

Exercise 10

a. Democrats lost their most conservative members, so overall Democratic membership is now more liberal.

b. Republicans gained more liberal-side members, so overall Republican membership is now more liberal.

c. Even though both groups individually shifted more liberal, it is also true that the House members that lost were replaced by individuals who were more conservative, so the overall House membership has become more conservative.

d. If the reporting is about overall votes in the House, the overall membership is more conservative and is likely to pass more conservative measures. If the reporting has to do with policies from only one party, then ideas suggested by each side likely will be more liberal than their party has put forth in the past. This conundrum is known as Simpson's Paradox.