

Mock FRQ #2

In a long-term study of 3462 randomly selected adults from Lausanne, Switzerland, researchers investigated the relationship between weekly napping frequency and whether or not a person experienced an event related to cardiovascular disease (CVD), such as a heart attack. The two-way table summarizes the data.

		Napping frequency per week				Total
		None	1-2	3-5	6-7	
CVD event status	Yes	93	12	22	28	155
	No	1921	655	389	342	3307
Total		2014	667	411	370	3462

- (a) In the sample, 4.5% of people experienced a CVD event. To what population can this result be generalized? Explain your answer.

Due to the random sampling, researchers can generalize this result to the population from which the sample was selected: all adults in Lausanne, Switzerland.

- (b) One person from those surveyed will be selected at random.

- (i) What is the probability that the person selected will be someone who reported taking at least 3 naps per week? You can leave your answer as an unreduced fraction.

$$\frac{411+370}{3462} = \frac{781}{3462} = 0.2256$$

- (ii) What is the probability that the person selected will be someone who has experienced a CVD event given that the person reported taking at least 3 naps per week? You can leave your answer as an unreduced fraction.

$$\frac{22+28}{781} = \frac{50}{781} = 0.064$$

Let p_F = the proportion of people who have experienced a CVD event among those in the population who would report frequent napping (at least 3 naps per week) and let p_I = the proportion of people who have experienced a CVD event among those in the population who would report infrequent napping (2 or fewer naps per week). A 95% confidence interval for $p_F - p_I$ is (0.0062, 0.0435).

- (c) One condition for constructing this confidence interval is that the number of people who experienced a CVD event and the number of people who did not experience a CVD event in each of the two groups is at least 10. Explain why it is necessary for this condition to be satisfied.

The condition is necessary because the formula for the confidence interval relies on the fact that the sampling distribution of $p_F - p_I$ is approximately normal. This will be the case if $n_F p_F$, $n_F(1 - p_F)$, $n_I p_I$, and $n_I(1 - p_I)$ are all at least 10.

- (d) Interpret the confidence interval.

We are 95% confident that the interval from 0.0062 to 0.0435 captures the difference in the proportion of people who have experienced a CVD event among those in the population who would report frequent napping and the proportion of people who have experienced a CVD event among those in the population who would report infrequent napping. In other words, we are 95% confident

that the proportion of people who have experienced a CVD event is between 0.62 percentage points higher and 4.35 percentage points higher for frequent nappers.

- (e) Even though all the values in the confidence interval are positive, a cause-and-effect relationship between napping and CVD cannot be established from this study due to possible confounding. Identify a potential confounding variable in this context. Justify your answer.

Age is a potential confounding variable, because it could be related to both napping frequency and CVD event status for the adults in the study. For example, older adults probably tend to nap more frequently and also tend to experience more CVD events than younger adults. If both of these are true, then the result of the study (that adults who nap more frequently tend to experience more CVD events than adults who nap less frequently) would have been found even if napping frequency has no effect on the likelihood of having a CVD event.