Using Applets to Explore Statistical Relationships

I. Napping and cardiovascular health   *Relationships between 2 categorical variables*

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 a major issue related to cardiovascular disease (CVD), such as a heart attack or stroke. The two-way table summarizes the data.

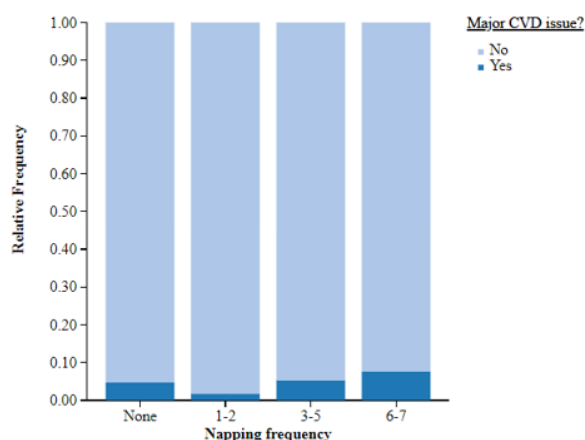|		Napping frequency per week				
		None	1-2	3-5	6-7	Total
Major CVD issue?	Yes	93	12	22	28	155
	No	1921	655	389	342	3307
	Total	2014	667	411	370	3462

Warm-up questions:

1. In the sample, 4.5% of people experienced a major CVD issue. To what population can this result be generalized? Explain your answer.

2. One person from those surveyed will be selected at random.
   (a) What is the probability that the person selected will be someone who reported taking at least 3 naps per week?

   (b) What is the probability that the person selected will be someone who has experienced a major CVD issue given that the person reported taking at least 3 naps per week?

3. Are the events "experienced a major CVD issue" and "reported taking at least 3 naps per week" independent for this sample of adults? Justify your answer.

∎∎∎

4. Go to www.stapplet.com. Then launch the *2 Categorical Variables* applet.
5. Choose to input the data as a Two-way table.
6. Enter "Napping frequency" as the explanatory variable and "Major CVD issue?" as the response variable.
7. Input the corresponding frequencies for each cell of the table. Use the + sign at the top right of the table to add columns.
8. Click the Begin analysis button. Segmented bar graphs for each category of the explanatory variable should be displayed, along with summary statistics.

Graph Distributions

Chart type: Segmented bar ▾

Major CVD issue?
- No
- Yes

Summary Statistics

		Napping frequency				
		None	1-2	3-5	6-7	Total
Major CVD issue?	Yes	93 (0.05)	12 (0.02)	22 (0.05)	28 (0.08)	155 (0.04)
	No	1921 (0.95)	655 (0.98)	389 (0.95)	342 (0.92)	3307 (0.96)
	Total	2014 (1)	667 (1)	411 (1)	370 (1)	3462 (1)

Proportions in table show the distribution of Major CVD issue? for each category of Napping frequency.

9. Do the data provide *some* evidence of an association between napping frequency and whether or not a person had experienced a major CVD issue in the population identified in Question 1? Explain your answer.

10. Scroll down to the Perform Inference section. Click on the Perform chi-square test for independence button. Record the test statistic and *P*-value.

11. Do the data provide *convincing* evidence of an association between napping frequency and whether or not a person had experienced a major CVD issue in the population identified in Question 1? Justify your answer.

12. Which napping frequency category is associated with the smallest risk of a major CVD issue? Explain.

Meadowfoam seed oil is used in making various skin care products. Researchers interested in maximizing the productivity of meadowfoam plants designed an experiment to investigate the effect of different light intensities on plant growth. The researchers planted 120 meadowfoam seedlings in individual pots, randomly assigned 10 pots to each of 12 trays, and put all the trays into a controlled enclosure. Two trays were then randomly assigned to each light intensity level (micromoles per square meter per second): 150, 300, 450, 600, 750, and 900. The number of flowers produced by each plant was recorded. Here are data on the average number of flowers per plant on each tray and the corresponding light intensity level.

Light intensity	150	150	300	300	450	450	600	600	750	750	900	900
Average number of flowers	62.3	77.4	55.3	54.2	49.6	61.9	39.4	45.7	31.3	44.9	36.8	41.9

(*Source:* M. Seddigh and G.D. Jolliff, "Light Intensity Effects on Meadowfoam Growth and Flowering," *Crop Science* 34 (1994): 497–503.)

1. Go to www.rossmanchance.com/applets/index2021.html. Choose Data Analysis→Least Squares Regression.
2. Open the Meadowfoam Experiment data file.
3. Copy and paste the data into the Enter Data panel.
4. Click the Use Data button. A scatterplot of the data should appear.
5. Describe the relationship between average number of flowers produced and light intensity for the meadowfoam plants in this experiment.

6. Identify and interpret the slope of the least-squares regression line in context.

7. Use these data to estimate the average number of flowers produced, *irrespective of light intensity.*

8. How much better does the least-squares regression line with x = light intensity do at predicting average number of flowers than using the mean $\overline{y}$ to predict average number of flowers *irrespective of light intensity*?

9. If we use the least-squares regression line to predict average number of flowers from light intensity, about how far off is a typical prediction?

10. Do the data give *some* evidence of a linear relationship between light intensity and average number of flowers produced for meadowfoam plants? Explain your answer.

Do these data provide convincing evidence of a linear relationship between light intensity and average number of flowers for meadowfoam plants like the ones in this experiment? Or is it believable that there is not actually a linear relationship between the two variables, and that the observed result ($b = -0.041$) occurred purely by the chance involved in the random assignment?

    To find out, let's simulate re-doing the random assignment under the assumption that there is no relationship between average number of flowers and light intensity. This amounts to assuming that each tray would have the same average number of flowers regardless of light intensity, and randomly reassigning the trays to the various light intensities, recording the slope b of the resulting sample regression line each time.

11. On the right side of the applet screen, click the box to Show Shuffle Options. Be sure that Cards is chosen under Select display. In the Choose statistic drop-down menu at the top right, select Slope.

12. Click the Shuffle Y-values button. What just happened?

13. Change the Number of Shuffles to 1000 and click Shuffle several times. What do you notice about the simulated distribution of the sample slopes?

14. What percent of dots on the graph have values at least as extreme (in either direction) as -0.0411? _____ (Use the Count shuffles Beyond feature under the graph to help answer this question.)

15. In other words, IF there is actually no linear relationship between light intensity and average number of flowers produced in the population of meadowfoam plants like the ones in this study, the estimated probability of obtaining a sample regression line with a slope at least as surprising (in either direction) as -0.0411 purely by chance due to random assignment is _____.

16. Based on the estimated probability in Question 15, do these data provide convincing evidence of a linear relationship between light intensity and average number of flowers produced in the population of meadowfoam plants like the ones in this study? Explain your answer.

17. In the left panel, click the radio button to show the Regression table. How does the estimated probability in Question 15 compare with the P-value in the computer output?

18. Now click the radio button to obtain a 95% Confidence Interval for Slope. Interpret this interval.