

*Inserted answers in italics; multiple choice answers in bold.*

1. A chi-square test of the relationship between personal perception of emotional health and marital status led to rejection of the null hypothesis, indicating that there is a relationship between these two variables. One conclusion that can be drawn is:
  - A. Marriage leads to better emotional health.
  - B. Better emotional health leads to marriage.
  - C. The more emotionally healthy someone is, the more likely they are to be married.
  - D. There are likely to be confounding variables related to both emotional health and marital status.**

**Questions 2 to 5:** A survey is done in which people are asked how often they exceed speed limits. The data are then categorized by age (Under 30, 30 and Over) and answer (Always, Not Always). The following computer output shows observed and expected counts, the chi-square statistic and the p-value.

Expected counts are printed below observed counts			
	Always	Not Always	Total
Under 30	100	100	200
	70.00	130.00	
30 and Over	40	160	200
	70.00	130.00	
Total	140	260	400
Chi-Sq =	12.857	+ 6.923	+
	12.857	+ 6.923	= 39.560
DF = 1,	P-Value =	0.000	

2. What is the relative risk of always exceeding the speed limit for people under 30 compared to always exceeding it for people 30 and over? Show your calculations.  
*Risk for under 30 =  $100/200 = .5$*   
*Risk for 30 and over =  $40/200 = .2$*   
*Relative risk =  $.5/.2 = 2.5$*

3. Write down the appropriate null hypothesis for this situation.

*There is no relationship between age group and whether or not one always speeds.*

4. Show how the Expected Count of 70 in the "Under 30, Always" cell was calculated.

$$(200)(140)/400 = 70$$

5. Make a conclusion about statistical significance and report it in the context of the situation.

*The p-value is 0.000, which is less than 0.05, so the relationship between age group and whether or not one always speeds is statistically significant.*

6. A chi-square test involves a set of counts called “expected counts.” What are the expected counts?
  - A. Hypothetical counts that would occur if the alternative hypothesis were true.
  - B. Hypothetical counts that would occur if the null hypothesis were true.**
  - C. The actual counts that did occur in the observed data.
  - D. The long-run counts that would be expected if the observed counts are representative.
  
7. Pick the choice that best completes the following sentence. If a relationship between two variables is called statistically significant, it means the investigators think the variables are
  - A. related in the population represented by the sample.**
  - B. not related in the population represented by the sample.
  - C. related in the sample due to chance alone.
  - D. very important.
  
8. Simpson's Paradox occurs when
  - A. No baseline risk is given, so it is not known whether or not a high relative risk has practical importance.
  - B. A confounding variable rather than the explanatory variable is responsible for a change in the response variable.
  - C. The direction of the relationship between two variables changes when the categories of a confounding variable are taken into account.**
  - D. The results of a test are statistically significant but are really due to chance.

**Questions 9 and 10:** A recent large randomized experiment generated substantial publicity when the researchers announced that the relative risk of breast cancer for post-menopausal women taking hormones compared to taking placebo was about 1.3.

9. Write a sentence explaining what relative risk means in this example, in words that someone with no training in statistics would understand.

*The risk of breast cancer for post-menopausal women taking hormones is 1.3 times the risk of breast cancer for women taking a placebo.*

10. The hormones in question have beneficial effects for relief of symptoms such as hot flashes. For women trying to make a decision about whether to take the hormones, why is it important to know the baseline risk of breast cancer in conjunction with the relative risk?

*Even though the relative risk is 1.3, if there is a very low baseline risk, then the increase in actual risk may be minimal. For instance, if the baseline risk is only 1/10,000 then this increase is minor. If the baseline risk is 1/100, then the increased risk is a big deal.*