Econ 57
Gary Smith
Spring 2023

Midterm Answers

1. Self-selection bias. Her readers were not a random sample of parents and those who responded were not a random sample of readers. Those who are unhappy with their children may have been more likely to respond. A scientific random sample asking the same question found that more than 90% answered *yes*.

2. Survivor bias. Many people with head injuries were not taken to hospitals because they were dead.

3. The interval widths are not equal, but the reason this figure is not a histogram is that the bar heights are counts—so the total area under the bars is not 1. [Shi, Z., Rui, H. & Whinston, A.B. (2014) "Content Sharing in a Social Broadcasting Environment: Evidence from Twitter", MIS Quarterly, 38(1): 123-142.]

4. Letting a red marble be a success
   a. binomial $P[x=3]=\binom{5}{3}0.6^3 0.4^2$
   b. with replacement $P[x=3]=\binom{5}{3}\frac{60}{100}\frac{59}{99}\frac{58}{98}\frac{40}{97}\frac{39}{96}$

   This can also be derived from

$$P[x=3]=\frac{\binom{60}{3}\binom{40}{2}}{\binom{100}{5}}=\frac{\frac{60(59)(58)}{3(2)(1)}\frac{40(39)}{2(1)}}{\frac{100(99)(98)(97)(96)}{5(4)(3)(2)1}}=10\frac{60(59)(58)40(39)}{100(99)(98)(97)(96)}$$

5. These calculations assume that the four risk factors are independent and they may not be.

6. This is a Bayes' Rule problem:

$$\begin{aligned}P[\text{human if }+]&=\frac{P[\text{human}]P[+\text{ if human}]}{P[\text{human}]P[+\text{ if human}]+P[\text{GPT-3}]P[+\text{ if GPT-3}]}\\&=\frac{0.9(0.1)}{0.9(0.1)+(0.1)(0.9)}\\&=0.50\end{aligned}$$

7. This is an example of Simpson's paradox. There were more women in occupations that had low unemployment rates.

8. a. No
   b. Yes
   c. No (e.g., with two independent coin flips, P[H1 or H2] $\neq$ P[H1]+ P[H2])
   d. Yes (A normal distribution is symmetrical with the median equal to the mean)
   e. No

9. Assuming that the guesses are independent, each with a 0.5 probability of being correct,
   a. $\mu = (-2)(0.5) = (1)(0.5) = -0.5$
   b. Guessing the answer to 1 question is best because every guess reduces the chances of a good score. With one guess, the probability of a positive score is 0.5.
      If Charlie guesses the answers to 2 or 3 questions, Charlie has to get all answers correct to get a positive score; those probabilities are less than 0.5. If Charlie guesses the answers to 4 questions, Charlie has to get 3 or 4 right to get a positive score.

$$P[3 \text{ or } 4 \text{ of } 4] = \binom{4}{3} 0.5^3 0.5^{4-3} + \binom{4}{4} 0.5^4 0.5^{4-4} = 4(0.5^4) + 0.5^4 = 5(0.5^4) = \frac{5}{16}$$

   and so on.

10. A time series graph is much more informative: