Week 1 - Data Science Math Assignments

Andrew Goldberg

7/5/15

Exercise 2.12: School absences. Data collected at elementary schools in DeKalb County, GA suggest that each year roughly 25% of students miss exactly one day of school, 15% miss 2 days, and 28% miss 3 or more days due to sickness.

(a) What is the probability that a student chosen at random doesn't miss any days of school due to sickness this year?

1- (.25+.15+.28) = **.32**

(b) What is the probability that a student chosen at random misses no more than one day?

.32+.25 = **.57**

(c) What is the probability that a student chosen at random misses at least one day?

.25+.15+.28 = **.68**

(d) If a parent has two kids at a DeKalb County elementary school, what is the probability that neither kid will miss any school? Note any assumption you must make to answer this question.

.32\*.32 = **.102**

Assumptions: Events are independent: neither sibling’s behavior is affected by other.

(e) If a parent has two kids at a DeKalb County elementary school, what is the probability that that both kids will miss some school, i.e. at least one day? Note any assumption you make.

.68 \* .68 = **.462**

Assumptions: Events are independent: neither sibling’s behavior is affected by other.

(f) If you made an assumption in part (d) or (e), do you think it was reasonable? If you didn't make any assumptions, double check your earlier answers.

The health of co-habiting siblings is very likely to be interrelated, so their absences would likely have some degree of dependence on each other, so the events would not really be independent

2.14 Weight and health coverage, Part I. The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey designed to identify risk factors in the adult population and report emerging health trends. The following table summarizes two variables for the respondents: weight status using body mass index (BMI) and health coverage, which describes whether each respondent had health insurance.66

Weight Status

Neither overweight Overweight Obese

nor obese (BMI < 25) (25 \_ BMI < 30) (BMI \_ 30) Total

Health Yes 134,801 141,699 107,301 383,801

Coverage No 15,098 15,327 14,412 44,837

Total 149,899 157,026 121,713 428,638

(a) If we draw one individual at random, what is the probability that the respondent is overweight and doesn't have health coverage?

15,327 / 428,638 = **.036**

(b) If we draw one individual at random, what is the probability that the respondent is overweight or doesn't have health coverage?

P(overweight) = 157026/428638 = .366

P(no coverage) = 44638/428638 = .104

P(overweight + no coverage) = .036

.366+.104 - .036 = **.435**

2.28 Socks in a drawer. In your sock drawer you have 4 blue, 5 gray, and 3 black socks. Half asleep one morning you grab 2 socks at random and put them on. Find the probability you end up wearing

(a) 2 blue socks

(4/12)\*(3/11) = **.090**

(b) no gray socks

(7/12)\*(6/11) = **.318**

(c) at least 1 black sock

1- ((9/12)\*(8/11)) = 1-.545 = **.455**

(d) a green sock

0

(e) matching socks

Blue = (4/12)\*(3/11) = .090

Gray = (5/12)\*(4/11) = .151

Black = (3/12)\*(2/11) = .045

Blue + Gray + Black = **.288**

2.30 Books on a bookshelf. The table below shows the distribution of books on a bookcase based on whether they are nonfiction or fiction and hardcover or paperback.

Format

Hardcover Paperback Total

Fiction 13 59 72

Nonfiction 15 8 23

Total 28 67 95

(a) Find the probability of drawing a hardcover book first then a paperback fiction book second

when drawing without replacement.

(28/58) \* (59/94) = **.184**

(b) Determine the probability of drawing a fiction book first and then a hardcover book second,

when drawing without replacement.

(72/95) \* ((28-(13/72))/94) = **.224**

(c) Calculate the probability of the scenario in part (b), except this time complete the calculations under the scenario where the first book is placed back on the bookcase before randomly drawing the second book.

(72/95) \* (28/95) = **.223**

(d) The final answers to parts (b) and (c) are very similar. Explain why this is the case.

Over a 95 book population, the small change of 1 book being replaced will have a minimal effect.