Data Science Math Week 1 Assignment

Data Science Math Week 1 Assignment

Exercise 2.12

P(miss 1) = 0.25, P(miss 2) = 0.15, P(miss 3 or more) = 0.28 due to sickness

1. P(miss 0) = 1- 0.25- 0.15- 0.28 = 0.32
2. P(miss no more than 1) = 0.32 + 0.25 = 0.57
3. P(miss at least 1) = 1 – 0.32 = 0.68
4. Assume that the attendance of each child is independent

P(2 children neither misses school) = 0.32 x 0.32 = 0.1024

1. Assume that the attendance of each child is independent

P(2 children both miss some school) = 0.68 x 0.68 =0.4624

1. I think the assumption of independence is reasonable on any given day (if the sickness is not contagious), however over the long term might not be accurate as elementary kids are notorious for passing germs. If the attendance is dependent then

P(child 2 miss 0|child 1 miss 0) = 0.32 x 0.32 / 0.32 = 0.32

P(child 2 miss some | child 1 miss some) = 0.68 \* 0.68 / 0.68 = 0.68

P(child 2 miss 0 | child 1 miss some) = 0.32 x 0.68 / 0.68 = 0.32 This is (and should be) the same as P(child 2 miss 0) independent of child 1 attendance

P(child 2 miss some | child 1 miss 0) = 0.68 x 0.32 / 0.32 = 0.68 This is (and should be) the same as P(child 2 miss some) independent of child 1 attendance

Exercise 2.14

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Weight Status |  |  |  |
|  |  | Neither overwt nor obese  BMI < 25 | Overweight  25 < BMI < 30 | Obese  BMI > 30 | Total |
| Health | Yes | 134801 | 141699 | 107301 | 383801 |
| Coverage | No | 15098 | 15327 | 14412 | 44837 |
|  | Total | 149899 | 157026 | 121713 | 428638 |

As Probabilities

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Weight Status |  |  |  |
|  |  | Neither overwt nor obese  BMI < 25 | Overweight  25 < BMI < 30 | Obese  BMI > 30 | Total |
| Health | Yes | 0.314 | 0.331 | 0.250 | 0.895 |
| Coverage | No | 0.035 | 0.036 | 0.034 | 0.105 |
|  | Total | 0.350\* | 0.366\* | 0.284 | 1.000 |

\*column total not exact match due to rounding

Overweight and obese are separate classes in the problems below

1. P(overweight AND no health coverage) = 0.036
2. P(overweight OR no health coverage) = 0.366 + 0.105 – 0.036 = 0.435

Exercise 2.28

4 blue, 5 gray, 3 black socks

Assume no replacement (you pick it, you wear it)

1. P(2 blue) = 5/12 x 4/11 = 20/132 = 5/33 = 0.15
2. P(no gray) = 7/12 x 6/11 =42/132 = 7/22 = 0.32
3. P(at least 1 black) = 3/12 x 11/11 = 33/132 = 1/4= 0.25
4. P(a green) = 0
5. P(matching socks) Not enough information. Just because socks are in the same color category does not mean they match. We have no information as to the number of matching socks in the drawer.

Exercise 2.30

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Format |  |  |
|  |  | Hardcover | Paperback | Total |
| Type | Fiction | 13 | 59 | 72 |
|  | Nonfiction | 15 | 8 | 23 |
|  | Total | 28 | 67 | 95 |

1. P(paperback fiction| hardcover) without replacement = 28 / 95 x 59 /94

= 0.185

b)

The probability of drawing a fiction book first then a hardcover book second without replacement are the pathways highlighted in red in the Venn diagram above.

Without replacement

P(hardcover|fiction) = (13/95)(15/94) + (13/95)(12/94) + (59/95)(13/94) + (59/95)(15/94)

=(195 + 156 + 767 + 885)/8930 = 2003/8930 = 0.2243

1. With replacement the probability of drawing a fiction book first then a hardcover book second is (72/95) x (28/95) = 0.223
2. The answers to b and c are similar because we are using a small sample size (2/95 = 2.1%) When the sample size is less than 10% observations are close to being independent even when sampling without replacement.

Optional challenge exercise

You could use the combn function.