Statistical Literacy

Problem Set 4

Rémi Viné
Due October 23rd, 2023

1. **Some computation.**

	1	2	3	4
x_i *(Swiss Francs)*	2	23	45	51
$P(X = x_i)$	0.3	0.1	0.2	0.4

Table 1: Probability distribution for a random variable X, the (fictitious) price to pay for train ticket Geneva - Bern, in CHF.

(a) Find the expected value, $E(X)$. Interpret.

(b) What is the median? Interpret and comment on the potential skewness of the distribution.

(c) Find the standard deviation, $\sigma(X)$. Interpret.

(d) Assume the SBB (Swiss train company) decides to stop the super cheap tickets (CHF 2) and proposes to those paying CHF 2 to now pay CHF 23. If there is no change in preferences (no substitution with other means of transportation so that all the CHF 2 buyers are shifted to the CHF 23 buyers), would the expected value increase or decrease? And what about the standard deviation? (You are not asked to compute the new statistics, you are simply asked to provide the intuition backing your responses)

2. **Probability.** The table, from IDMC 2022 report, indicates the number of internally displaced young people by region (stock). Probabilities are asked assuming one could select a person out of the universe of all internally displaced young people. You may use the given acronyms to facilitate the notation.

	0-4yo	5-14yo	15-24 yo
Sub-Saharan Africa (SSA)	4.4	7.3	5.5
Middle East and North Africa (MENA)	1.5	2.7	5
South Asia (SA)	0.9	1.8	1.5
Americas (A)	0.5	1	1.1
Europe and Central Asia (ECA)	0.2	0.5	0.4
East Asia and Pacific (EAP)	0.2	0.5	0.4

Table 2: Age groups and regions of origin of internally displaced young people (in million)

(a) What is the probability for an internally displaced young person to come from the Americas?

(b) What is the probability for an internally displaced young person to come from South Asia and to be between 5 and 14 years old?

(c) What is the probability for an internally displaced young person to be aged 15-24 years old or come from "East Asia and Pacific"?

(d) Given that the young internally displaced is aged 0-4 years old, what is the probability that this person comes from Sub-Saharan Africa?

(e) Are being 0-4 years old and coming from Sub-Saharan Africa independent events? Justify and interpret the answer to the question.

3. **More on probability.** ILO reports modern slavery around the world. It is constructed around a taxonomy that splits modern slavery into two categories: (i) forced labor, (ii) forced marriage. Table 3 displays some frequencies taken from a 2022 ILO report. The worldwide share of individuals under the yoke of modern slavery is 6.4‰. This accounts for 49'570'000 individuals. Answer each of the following questions, if possible, if not, simply indicate that there is not enough information to find the answer. In your computation, stick using fractions, using decimal format (with four decimals) only for the results.

(frequencies in thousands of people)	Forced labor (FL)	Forced Marriage (FM)
Male (M)	15779	·
Female (F)	11798	·
Adults (A)	·	·
Children (C)	3314	8973
High income (Hi)	5384	1865
Upper middle income (Umi)	8965	·
Lower middle income (Lmi)	8916	·
Low income (Li)	4312	2260

Table 3: Global estimate of modern slavery - Forced labor and forced marriage, 2022

(a) Indicate what is the probability to be a male (worldwide) and to suffer from forced labor (in ‰, and assuming events "to be a male" and "to suffer from modern slavery" are independent events).

(b) What is the probability for an individual suffering from modern slavery to be female and in a forced marriage situation?

(c) Among individuals facing modern slavery, what is the probability that the individual is a child?

(d) Among individuals facing modern slavery, what is the probability that the individual is a child doing forced labor?

(e) Given that the individual in modern slavery is in forced marriage, what is the probability that the person is in a high income country?

(f) Are "age group" (A, C) and "type of modern slavery" (FL, FM) independent?

4. **Multiple choice.**

   (a) Based on online bets, for the last two matches of the season, the probability for Lausanne HC (the ice hockey team) to win against Genève-Servette HC is 0.4, the probability for Lausanne HC to win against HC Ajoie is 0.8, and the probability to win both games is 0.3. What is the probability to win at least one match?

   A 1.3

   B 0.9

   C 0.8

   D 0.6

   (b) What is the probability to end the season with one loss and one win?

   A 0.4

   B 0.9

   C 0.6

   D −0.1

   (c) Throw two dice, what is the probability to obtain *at least one "5"?*

   A 0.26

   B 0.13

   C 0.31

   D 0.54

   (d) *Difficult!* In the context of a test for infection, we know that an infected person will get a positive test result 90% of the time and an uninfected person will get a negative test result 95% of the time. We also know that about 30% of the individuals are infected. We want to know the probability that a person is infected given that she tests positive.

   A 0.89

   B 0.73

   C 0.56

   D 0.98