Statistical Literacy

Problem Set 2

Rémi Viné
Due October 9th, 2023

1. **Frequency tables and visualization of people attending a show**

Country	Frequency
India	4
Nigeria	2
Germany	2
China	2
Mexico	4
USA	6

Table 1: People attending a show

Table 1 shows the number of people attending a show by country of origin.

(a) What is the total number of people attending the show?

(b) What group of people, by country of origin is the most frequent? What is its share?

(c) What is the type of variable "Country"?

(d) What could be an appropriate data visualization? *Optional: you can try to do it using Stata*

(e) If possible, provide the different centrality indicators and briefly interpret them.

(f) Assume now that more details are available (table 2), in particular the type of show attended. What would be a suited data visualization? *(Optional: you can try to make one on Stata.)*

Country	Cinema	Concert	Sport
India	3	0	1
Nigeria	0	1	1
Germany	2	0	0
China	1	1	0
Mexico	1	2	1
USA	4	1	1

Table 2: People attending a show, by type of show

2. **Frequencies and visualizations on the variable daily hours worked.** Table 3 shows the 20 workers of a company and informs on the number of hours worked in the day.

Observation number	Daily hours worked
1	4
2	10
3	9
4	12
5	6
6	8
7	8
8	7
9	8
10	10
11	8
12	9
13	8
14	10
15	8
16	8
17	10
18	8
19	8
20	9

Table 3: Workers in a given place

(a) What is the type of variable "Daily hours worked"?

(b) Produce a new table grouping individuals by the number of hours worked.

(c) What is the most frequent number of hours worked? What is the share of em-

ployees working this amount of time?

(d) What is the percentage of students working less than 8 hours?

(e) What is the percentage of students working more than 8 hours?

(f) What is the percentage of students working between 6 and 9 hours (6 and 9 hours included)?

(g) What type of data visualization is appropriate? *(Optional: you can try to make one on Stata.)*

(h) If possible, provide the different centrality indicators.

3. **Frequencies and visualizations on height variable.** Table 4 groups individuals by their height in centimeters.

Frequency	Height (in centimeters) intervals
3	$[100 - 140)$
76	$[140 - 155)$
123	$[155 - 165)$
286	$[165 - 175)$
141	$[175 - 185)$
31	$[185 - 195)$
2	$[195 - 210)$

Table 4: Height of students

(a) What is the type of variable of the grouped heights?

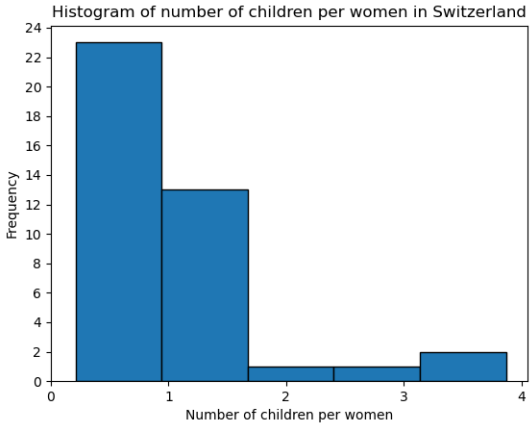
(b) Compute the cumulative frequency, the relative frequency, and find the midpoints of each intervals.

(c) What is the most relevant data visualization for this variable? *(Optional: you can try to make one on Stata.)*

(d) If possible, provide the different centrality indicators.

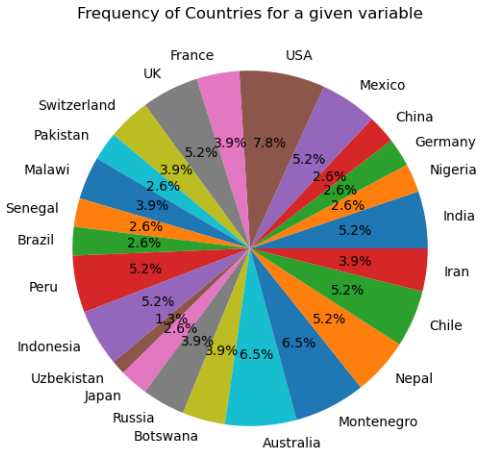
4. **Read charts**

(a) Can you inform on the number of observations in the sample used to construct the following histogram?

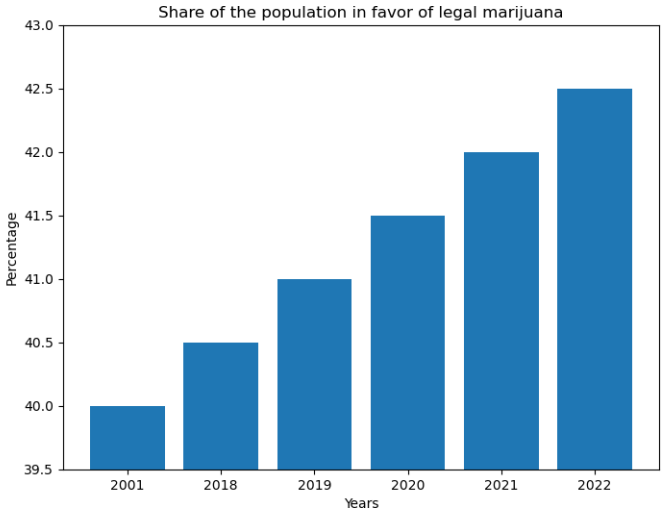


(b) Do you suspect, visually, a skewness? Do you expect a median larger than the average or lower than the average?

(c) Looking at this pie chart, do you consider it as appropriate? If not, why and what would you change?



(d) Looking at this bar chart, do you consider it as appropriate? If not, why, what do you think is the agenda of the person making such graph, and what would you change?


Share of the population in favor of legal marijuana
43.0
42.5
42.0
41.5
41.0
40.5
40.0
39.5
Percentage
2001
2018
2019
2020
2021
2022
Years