

Homework: Survey and Lottery

Statistical models

Lottery

The game 'Cash' is a lottery in France where each ticket costs 1 euro, offering the chance to win more than the ticket price. With 1,250,000 tickets considered, the odds of winning are presented in Table 1.

Table 1: Number of tickets and prize amounts for the 'Cash' lottery.

Number of tickets	Prize amount
1	25000
1	5000
2	1000
3	100
4	60
125	50
1900	20
3750	15
25500	10
23500	5
37500	2
214000	1
943714	0

1. Compute the expectation of the game and interpret your result.
2. Compute the variance of the game.
3. If you buy 100 tickets, what is the approximate probability that your gains or losses will exceed 0 euros? Hint: Use the central limit theorem to solve this.
4. How would you adjust the number of tickets for each prize amount to ensure that the probability of your gains or losses being more than 0 euros is about 50%?

Survey

The dataset 'Survey_netflix.xlsx' contains students' preferences between Netflix and Disney+. A value of 1 indicates a preference for Netflix. The purpose of this exercise is to review the concepts from the first chapter by determining whether both streaming platforms are equally preferred.

1. Compute the average and the empirical variance of the 'Netflix' variable.
2. Let us assume that the 'Netflix' variable is generated from a sequence of independent Bernoulli distributions with probability parameter p . The parameter p represents the population preference probability for watching movies on Netflix. Let us consider the following hypothesis:

$$H_0 : p = 0.8 \text{ versus } H_A : p > 0.8$$

In your own words, explain what these hypotheses are testing.

3. Use the binomial distribution to compute the p-value of the one-sided hypothesis.
4. Consider again the one-sided test and assume that the central limit theorem applies. Use the Normal distribution to compute the p-value of the one-sided hypothesis.
5. Discuss the two p-values computed in the two last questions. Do you reach the same conclusion about the hypothesis ?
6. Assume we want to find out whether Netflix is more preferred than Disney+ in Italy. How many individuals should we survey to ensure a 5% margin of error at a 95% confidence level? In other words, how many people should we survey to ensure the confidence interval's upper and lower bounds deviate by 5% from the estimated probability?