Ngoc Phan
February 23, 2020

Lending Club
2018 Loan Data
Inferential Statistics

Purpose

Apply inferential statistics to answer the following questions:

- What is the return on investment (ROI)?
- What is the loss?
- How many months the borrowers are on the loan?
- What is the expected number of defaults for a given sample?

Variables in-study

- **Return on Investment (ROI)**
  - Formula: total loan payment – loan amount
  - Examine ROI by loan status (fully paid vs. default) and by loan term (36 months vs. 60 months).
- **Months in Loan**
  - Formula: last payment date – loan issue date
  - Examine the length of loan in months.

Statistical Methods

- Apply t-test to see if there is a significant difference on the mean between 2 subgroups, e.g. loan status and loan term.
- Apply binomial distribution to estimate the expected number of defaults for a given sample.

Findings

- **Return on Investment (ROI)**
  For fully paid loans, we are 95% confident that the average ROI is:
  - between 990 and 1015 dollars, regardless of loan term.
  - between 756 and 770 dollars for 36 months loan.
  - between 1776 and 1820 dollars for 60 months loan.
- **Loss**
  For default loans, we are 95% confident that the average loss is:
  - between 13,490 and 13,744 dollars, regardless of loan term.
  - between 10,744 and 11,040 dollars for 36 months loan.
  - between 17,810 and 18,185 dollars for 60 months loan.
- **Months in Loan**

Most borrowers stay on the loan for an average of 7 months.

- **Estimate the expected number of defaults and loss**
  The table below shows the expected number of defaults and loss based on the number of invested loans. The data generated is based on the following assumptions:
  - The lender invests $250 in each loan.
  - The lender loses all money if the loan is default.
  - There is no other hidden fees or costs.

  Variables:
  - **no_loans**: number of loans to be invested in
  - **expected_no_defaults**: the expected number of defaults obtained from binomial distribution, which is the mean obtained from the distribution.
  - **tot_inv**: total amount invested = amount invested in each loan * no_loans
  - **expected_loss**: expected loss = amount invested in each loan * expected_no_defaults

	no_loans	expected_no_defaults	tot_inv	expected_loss
0	50	9.0	12500.0	2250.0
1	100	19.0	25000.0	4750.0
2	150	28.0	37500.0	7000.0
3	200	37.0	50000.0	9250.0
4	250	47.0	62500.0	11750.0
5	300	56.0	75000.0	14000.0
6	350	66.0	87500.0	16500.0
7	400	75.0	100000.0	18750.0
8	450	84.0	112500.0	21000.0
9	500	94.0	125000.0	23500.0
10	550	103.0	137500.0	25750.0
11	600	112.0	150000.0	28000.0
12	650	122.0	162500.0	30500.0
13	700	131.0	175000.0	32750.0
14	750	141.0	187500.0	35250.0
15	800	150.0	200000.0	37500.0
16	850	159.0	212500.0	39750.0
17	900	169.0	225000.0	42250.0
18	950	178.0	237500.0	44500.0
19	1000	187.0	250000.0	46750.0