Movie Rating

Mateusz Majak, Przemysław Mazurek

9 06 2021

Main purpose

- Reproduce old project
- Check consistency of findings on a new dataset

Reproduce old project

- new software (R vs. Stata)
- lack of documentation about data preparation -> we tried to rely on report but information attached was too vague
- not able to reproduce the exact results

Data preparation

- originaly performed in Excel
- this time we focused on obtaining the same data rather than on reasonable decisions

Preparation based on report description...

...did not result in the same outcome

Old:

VARIABLES	N	mean	sd	min	max
vote_average	3112	6.334	0.837	2.300	8.500
runtime	3112	111.0	20.96	63	338
popularity	3112	30.03	36.46	0.179	875.6
years_old	3112	16.99	12.74	3	103
revenue	3112	1.256e+08	1.884e+08	6399	2.788e+09
budget	3112	4.187e+07	4.471e+07	7000	3.800e+08
vote_count	3112	1013	1428	20	13752

New:

```
================================================================
Statistic      N      Mean          St. Dev.       Min       Max
----------------------------------------------------------------
vote_average 4,800    6.094            1.188       0.000    10.000
runtime      4,800  106.898           22.562       0         338
popularity   4,800   21.506           31.822       0.0004   875.581
years_old    4,800   15.571           12.418       1         102
revenue      4,800 82,312,052.000 162,895,005.000 0     2,787,965,087
budget       4,800 29,060,068.000  40,730,295.000 0       380,000,000
vote_count   4,800  690.647        1,234.852       0        13,752
----------------------------------------------------------------
```

Trimming values to the range from old project

Old:

VARIABLES	N	mean	sd	min	max
vote_average	3112	6.334	0.837	2.300	8.500
runtime	3112	111.0	20.96	63	338
popularity	3112	30.03	36.46	0.179	875.6
years_old	3112	16.99	12.74	3	103
revenue	3112	1.256e+08	1.884e+08	6399	2.788e+09
budget	3112	4.187e+07	4.471e+07	7000	3.800e+08
vote_count	3112	1013	1428	20	13752

New:

```
==================================================================
Statistic      N       Mean         St. Dev.       Min      Max
------------------------------------------------------------------
vote_average  3,041      6.338          0.840       2.300    8.500
runtime       3,041    111.023         21.078         63      338
popularity    3,041     29.516         35.383       0.212   875.581
years_old     3,041     16.317         12.698          3      102
revenue       3,041 123,835,650.000 186,283,144.000 6,399 2,787,965,087
budget        3,041 41,353,351.000 44,018,676.000   7,000 380,000,000
vote_count    3,041    998.206      1,412.578         20   13,752
------------------------------------------------------------------
```
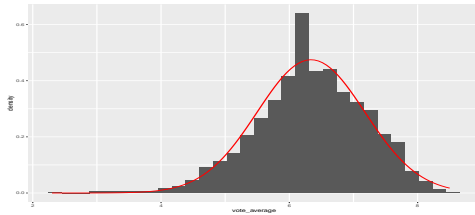
Some differences

Old:
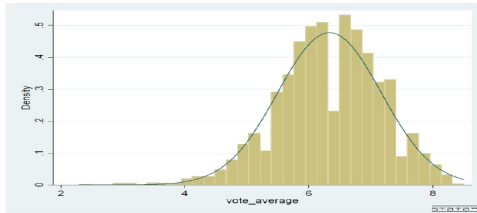
New:

Results

- quite similar at first glance
- passed RESET specification test originally
- failed RESET specification test at 5% of significance level
- error isn't normally distributed (previously it probably was)