

# Индекс стабильности популяции (Population stability Index)

Синонимы: PSI

Разделы: [Метрики](#)

Индекс стабильности популяции в [бизнес-аналитике](#) является широко распространенной [метрикой](#) для [мониторинга](#) актуальности аналитических моделей. Он отражает разницу между [обучающей выборкой](#) и данными, на которых модель используется.

Индекс стабильности популяции интерпретируется следующим образом:

- PSI менее 10% показывает отсутствие значимого изменения в текущей выборке;
- PSI в диапазоне от 10 до 25% свидетельствует о незначительном изменении, которые необходимо исследовать;
- PSI более 25% говорит о значительном смещении популяции и требуется перестроение модели.

Наиболее часто индекс стабильности популяции используется при мониторинге моделей в [кредитном скоринге](#), однако он может применяться и в других приложениях, где используются модели, строящиеся «от данных». Использование таких моделей в [системах поддержки принятия решений](#) всегда сопряжено с определенным риском, который связан с неблагоприятными последствиями решений принятых из-за некорректных результатов работы модели.

Именно поэтому необходимо оценивать актуальность и [репрезентативность](#) не только обучающей выборки, но и текущей выборки, на которой модель используется, чтобы убедиться что она соответствует данным. И если соответствие нарушено, то необходимо скорректировать или заменять модель. Для этих целей и служит индекс стабильности популяции.

Период относительно которого производится оценка стабильности популяции называется базовым, а период для которого производится оценка — целевым. Для примера рассмотрим распределения обучающей (базовое) и текущей (целевое) выборок заемщиков по диапазонам скорингового балла (см. таблицу).

Score	Целевое (t), %	Базовое (b), %	Ц- Б	Ц/Б	$\ln(\text{Ц}/\text{Б})$	Индекс
0-169	7	8	-1	0,875	-0,1335	0,0013

Score	Целевое (t), %	Базовое (b), %	Ц- Б	Ц/Б	ln(Ц/ Б)	Индекс
170- 179	8	10	-2	0,800	-0,2231	0,0045
180- 189	7	9	-2	0,778	-0,2513	0,0050
190- 199	9	13	-4	0,692	-0,3677	0,0147
200- 209	11	11	0	1,000	0,0000	0,0000
210- 219	11	10	1	1,100	0,0953	0,0010
220- 229	10	9	1	1,111	0,1054	0,0011
230- 239	12	10	2	1,200	0,1823	0,0036
240- 249	11	11	0	1,000	0,0000	0,0000
250+	14	9	5	1,556	0,4418	0,0221
					PSI	0,0533

Пусть базовая популяция была получена в 2007 году, а целевая в 2017. Требуется оценить, значимо ли различие между ними по методу PSI. Индекс стабильности популяции учитывает разницу между фактическим и ожидаемым распределениями и рассчитывается следующим образом:

$$PSI = \sum_{i=1}^N (y_{ti} - y_{bi})(\ln(y_{ti}) - \ln(y_{bi})) = \sum_{i=1}^N (y_{ti} - y_{bi}) \ln \frac{y_{ti}}{y_{bi}},$$

где  $y_{ti}$  — значение показателя целевого периода, попавшая в  $i$ -й интервал,  $y_{bi}$  — значение показателя базового периода. Тогда для первой строки таблицы, можно вычислить:  
 $(0,07 - 0,08) \cdot \ln(0,07/0,08) = 0,0013$ .

Индекс стабильности популяции является мерой изменения популяции за определенный период времени. Он показывает насколько «устаревает» модель, например скоринговая карта, за этот период.

Индекс стабильности популяции является одним из основных показателей точности и актуальности скоринговой карты в решении Loginom Scorecard Modeler. Подробнее в выступлении «Построение скоринговых карт в Loginom».