

УТВЕРЖДЕНЫ

приказом Росстата

от 03.04.2018 № 155

Методологические положения и методы импутации и авторедактирования данных по формам федерального статистического наблюдения № П-1 «Сведения о производстве и отгрузке товаров и услуг» и № П-5(м) «Основные сведения о деятельности организации»

Настоящие методологические положения описывают методы импутации и авторедактирования первичных статистических данных по формам федерального статистического наблюдения № П-1 «Сведения о производстве и отгрузке товаров и услуг» и П-5(м) «Основные сведения о деятельности организации» (далее форма № П-1 и № П-5(м) соответственно), позволяющие импутировать статистические данные, не вошедшие в своды в результате не предоставления респондентами отчетов, или скорректировать предоставленные недостоверные данные.

Процедуры импутации и авторедактирования применяются в отношении организаций, не относящихся к субъектам малого предпринимательства, деятельность которых не оказывает существенного влияния на сводные итоги по показателям вышеуказанных форм при их значительном количестве.

I. Общие положения

Процедуры импутации и авторедактирования применяются к первичным статистическим данным, предоставленным респондентами, и только для формирования сводных итогов на федеральном уровне.

Для данной работы единицей импутации является полный отчет, а не один или несколько показателей. Импутировать отдельные показатели невозможно ввиду того, что должны учитываться существующие статистические взаимосвязи между всеми показателями формы.

Если отчет или отдельный показатель предприятия хотя бы один раз подвергался процедурам импутации или авторедактирования, то все данные предприятия по этому показателю не могут участвовать в выборке для вычисления расчетных значений в будущем.

При предоставлении респондентом откорректированного отчета с уточненным показателем, который ранее был импутирован системой, проводится уточнение ранее предоставленных отчетных данных и замена импутированных показателей на отчетные.

Результаты работы описанных методов импутации и авторедактирования не должны оказывать существенного влияния на полученные сводные итоги. Вследствие чего, процедурам импутации или авторедактирования не подвергаются отчеты статистически значимых организаций.

В связи с тем, что поступление первичной статистической информации по различным видам экономической деятельности и видам продукции может иметь неравномерное распределение в рамках всего года, виды деятельности (ОКВЭД2) и виды продукции (ОКПД2) в рамках региона делятся на три категории:

регулярные – виды экономической деятельности и виды продукции, по которым первичная статистическая информация предоставляется в каждом месяце в течение года, и при этом отношение максимального суммарного значения показателя в регионе к минимальному суммарному значению показателя в регионе составляет не более 500%);

нерегулярные – виды экономической деятельности и виды продукции, по которым статистические данные могут быть представлены организациям региона в одном или нескольких месяцах года, и при этом в этих периодах наблюдаются случайные скачки суммарного значения показателя по региону по отношению к минимальному суммарному значению показателя в течение года более 500%);

сезонные – нерегулярные, но их появление характерно для конкретных периодов года (в определенные месяцы года имеются скачки значений показателя относительно суммарного минимального значения более 500%).

Проведение процедуры импутации для предприятий осуществляется:

с регулярными видами экономической деятельности и видами продукции – методом прогнозирования по динамике и методом донора;

с сезонными видами экономической деятельности и видами продукции – только в соответствии с методом донора и только для месяцев, когда наблюдаются скачки суммарного минимального значений показателя по региону более 500%;

с нерегулярными видами экономической деятельности и видами продукции – не осуществляется.

Импутация не применяется, если непредоставление отчетности респондентом связано с прекращением его деятельности. Прекращение деятельности организации определяется по критерию ведения деятельности предприятием и, как правило, является экспертной оценкой экономиста.

В основе алгоритмов импутации и авторедактирования лежат методы прогнозирования по динамике и донора.

Метод прогнозирования по динамике применяется при наличии данных по предприятию за последние 12 месяцев, позволяющих получить не менее 3 прогнозируемых значений показателя.

В случае, если для прогнозирования по динамике предприятия недостаточно данных, возможно применение метода донора.

II. Критерий статистической значимости

Статистически значимое предприятие – предприятие, значение хотя бы одного показателя которого за последний год вносит критически весомый вклад в статистику региона по показателю в разрезе вида продукции или вида деятельности. Значение показателя предприятия составляет X% – от суммарного значения показателя по региону в вышеуказанном разрезе. При этом рассматривается каждый из 12 последних месяцев. Величина X% варьируется и зависит от:

территории;
вида продукции (ОКПД2)/вида деятельности предприятия (ОКВЭД2).

Предприятие не может быть одновременно по одному показателю значимым (не импутироваться) и незначимым по другому (подвергаться процедуре импутации). Не является статистически значимым то предприятие, которое ни по одному из показателей не попало в число значимых.

Перечень показателей, на основе которых определяется статистическая значимость организаций для форм П-1 и П-5 (м):

отгружено товаров собственного производства, выполнено работ и услуг собственными силами (без НДС и акцизов) (по основным ОКВЭД2, раздел 1, строка 1);

отгружено товаров собственного производства, выполнено работ и услуг собственными силами (без НДС и акцизов) (по "чистым" ОКВЭД2 - из раздела 2, строки 21);

продано товаров несобственного производства;

оборот розничной торговли;

оборот оптовой торговли;

перевезено грузов грузовыми автомобилями;

произведено важнейших видов промышленной продукции;

розничная продажа основных товаров;

оптовая продажа отдельных товаров организациями оптовой торговли.

Подход, применяемый для определения и визуализации статистически значимых значений показателей респондентов, представлен на рис.1.

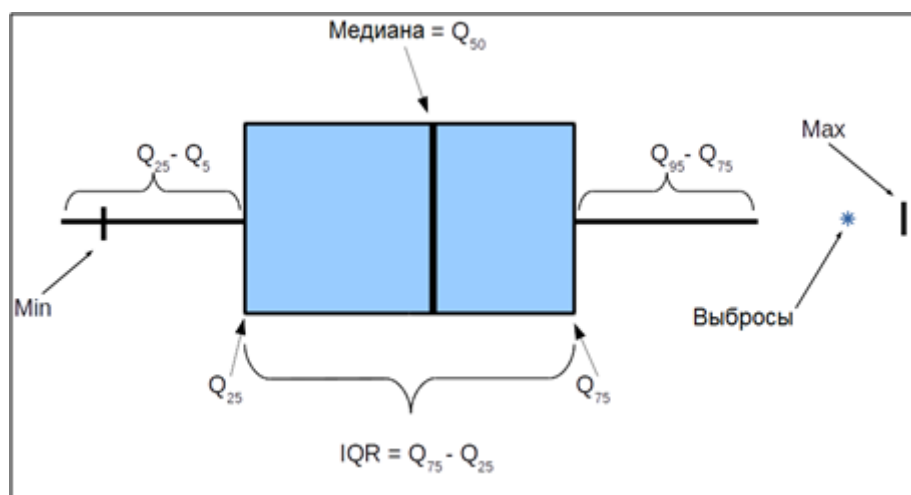


Рис.1

IQR – межквартильный интервал между 1 и 3-м квартилем;
 Max – максимальное значение показателя;
 Min – минимальное значение показателя.

Границами ящика служат первый (Q_{25}) и третий (Q_{75}) квартили, линия в середине ящика — медиана (Q_{50}).

Начиная с 95-го перцентиля и до максимального значения необходимо определить долю организаций в общем объеме по показателю, значения которых попали в этот интервал, и рассчитать разницу между этими долями.

При первом обнаружении разницы между долями организаций более 30% на данном интервале считать организацию с большей долей и все последующие за ней статистически значимыми.

III. Критерий ведения деятельности предприятием

После выявления незаполненных показателей на основании данных прошлых периодов по этому показателю определяется тренд к снижению его значения. Если такой тренд имеет место, и в текущем периоде показатель отсутствует, то импутировать такой показатель не следует.

Перечень показателей, на основе которых определяется критерий ведения деятельности предприятием для форм П-1 и П-5(м):

отгружено товаров собственного производства, выполнено работ и услуг собственными силами (без НДС и акцизов) (по основным ОКВЭД2, раздел 1, строка 1);

отгружено товаров собственного производства, выполнено работ и услуг собственными силами (без НДС и акцизов) (по "чистым" ОКВЭД2 - из раздела 2, строки 21);

продано товаров несобственного производства;

остатки важнейших видов промышленной продукции;

товарные запасы в организациях (предприятиях) розничной торговли;

запасы отдельных товаров в организациях оптовой торговли.

Продолжение ведения деятельности предприятия по конкретному показателю и коду ОКПД2/ОКВЭД2 определяется наличием /отсутствием тренда к полному прекращению деятельности по данному показателю (Р) и коду ОКПД2/ОКВЭД2, определяемого по формуле:

$$P_k = P_{k-1} + \frac{\sum_{i=0}^5 (P_{k-i} - P_{k-1-i})}{5}$$

где:

- k - отчетный месяц,
- (k-1) - предыдущий месяц,
- 5 - количество месяцев,
- $\frac{\sum_{i=0}^5 (P_{k-i} - P_{k-1-i})}{5}$ – средний прирост /убывание показателя за отчетный период. После расчета этого показателя анализируется возможность реального сведения его к нулю в текущем месяце.

Если значение показателя (P_k), полученного по вышеприведенной формуле, оказывается ≤ 0 , тренд к полному прекращению деятельности однозначно присутствует, возможность импутации показателя исключается.

Если значение показателя (P_k), оказалось > 0 , но при этом $\frac{\sum_{i=0}^5 (P_{k-i} - P_{k-1-i})}{5} \leq 0$, то отсутствие тренда к полному прекращению деятельности по данному показателю не исключается, и окончательное определение критерия по данной организации осуществляется экономистом.

Если показатели производства и отгрузки имеют тренд к снижению, но при этом показатели остатков на текущий период сохраняются положительными, следует считать, что такой респондент продолжает вести деятельность.

IV. Прогнозирование по динамике предприятия

Прогнозирование по динамике предприятия осуществляется по формуле, соответствующей методике экспоненциального сглаживания:

$$\hat{Y}_{t+1} = k * Y_t + (1-k) * \hat{Y}_t = k * \sum_{i=0}^{t-1} (1-k)^i Y_{t-i} + (1-k)^t * Y_1$$

где:

- \hat{Y}_{t+1} – прогноз на период t+1;
- Y_t – данные для прогноза за текущий период t;
- k - коэффициент сглаживания ряда, k задается экспертно и находится в диапазоне от 0 до 1, $0 < k < 1$
- \hat{Y}_t – значение прогноза на текущий период t. Причем в первый период $\hat{Y}_1 = Y_1$. Т.е. \hat{Y}_t в первый период равны значению показателя за этот период.

V. Метод донора

Суть метода заключается в формировании набора респондентов, которые станут донорами для текущего респондента в текущем периоде.

Период, в котором был сдан последний отчет, называется показательным.

Доноры выбираются по совпадению кодов основного ОКВЭД2, типа предприятия и признака «численность предприятия» (до 15 человек или более 15 человек) в регионе (в текущем и показательном периоде). Для показателей 2 и 5 разделов формы № П-1 дополнительным требованием является совпадение кодов ОКПД2 и фактических кодов ОКВЭД2 (для раздела 2: по каждому ОКВЭД2 выбирается перечень доноров, которые предоставили информация по данному коду ОКВЭД2 в показательном периоде; для раздела 5: по каждому ОКПД2 выбирается перечень доноров, которые предоставили информацию по данному коду ОКПД2 в показательном периоде).

Для осуществления процедуры импутации на выбранной территории (субъект или несколько соседних субъектов) искомый код ОКПД2/ОКВЭД2 должен быть представлен двумя и более предприятиями.

В случае, если для одного или нескольких показателей респондента не было найдено доноров в данном регионе, поиск доноров должен осуществляться во всех соседних регионах. Если доноров не было найдено в соседних регионах, поиск должен осуществляться по всем регионам.

Если для одного или нескольких показателей респондента не было найдено доноров по всем регионам, поиск доноров необходимо осуществлять по совпадению первых 4 знаков хозяйственного кода ОКВЭД2.

Для каждого донора по каждому импортируемому показателю определяется коэффициент влияния по формуле:

$$\delta_i = \frac{P^{k-1}}{P_i^{k-1}}$$

где δ_i – коэффициент влияния i -ого донора, P^{k-1} – значение показателя респондента, данные которого должны быть импутированы, из отчета показательного периода, P_i^{k-1} – значение показателя i -ого донора из отчета показательного периода. Коэффициент влияния показывает отношение данных респондента и данных донора.

Для респондента должен вычисляться коэффициент роста (снижения) производства за последние 3 месяца до показательного периода:

$$t = \frac{\frac{P^{k-1}}{P^{k-2}} + \frac{P^{k-2}}{P^{k-3}} + \frac{P^{k-3}}{P^{k-4}}}{3}$$

где P^{k-i} – значение показателя за периоды, предшествующие показательному. Коэффициент роста (снижения) производства показывает

тенденции предприятия к росту или снижению производства и продаж за последние 3 месяца.

Импутированное значение показателя за текущий период для респондента определяется по формуле:

$$P^k = t \frac{1}{n} \sum_{i=1}^n \delta_i P_i^k = t \frac{1}{n} \sum_{i=1}^n \frac{P_i^{k-1}}{P_i^{k-1}} P_i^k$$

где n – количество доноров, P_i^k – значение показателя i -ого донора за текущий отчетный период. Формула учитывает тенденцию предприятия к росту или снижению производства в текущем году и влияние каждого донора на итоговое значение показателя.
