

ИСПОЛЬЗОВАНИЕ ПРЕДИКЦИОННЫХ ИНТЕРВАЛОВ ПРИ КОНТРОЛЕ КАЧЕСТВА ПО АЛЬТЕРНАТИВНОМУ ПРИЗНАКУ

Борбаць Н.М., Школина Т.В.

Брянский государственный технический университет, Брянск

Ключевые слова: предикционный интервал, контроль качества, число несоответствий, число несоответствующих единиц, язык R.

Аннотация. Предикционные интервалы, содержащие с заданной степенью доверия одно, все или как минимум k наблюдений из их заданного числа в будущей выборке, зачастую представляют больший интерес при контроле качества, чем другие виды интервалов, особенно при малом числе единиц продукции. Тем не менее, вопросы построения подобных интервалов для результатов контроля по альтернативному признаку крайне скудно освещены в литературе. Авторами предлагается функция на языке R, позволяющая легко находить границы двустороннего предикционного интервала и одностороннего предикционного предела для числа несоответствующих единиц продукции.

USING PREDICTION INTERVALS IN QUALITY CONTROL BY ATTRIBUTES

Borbats N.M., Shkolina T.V.

Bryansk State Technical University, Bryansk

Keywords: prediction interval, quality control, number of nonconformities, number of nonconforming units, R language.

Abstract. Prediction intervals containing, with a specified degree of confidence, one, all, or at least k observations from a given number of them in a future sample are often of greater interest in quality control than other types of intervals, especially with a small number of units of production. Nevertheless, the issues of constructing such intervals for control results by attributes are extremely poorly covered in the literature. The authors propose a function in the R language that allows you to easily find the boundaries of the two-sided prediction interval and one-sided prediction limit for the number of nonconforming units of production.

В практике контроля качества применяются различные инструменты и методы математической статистики. Одним из таких инструментов являются различные статистические интервалы. Наиболее известными и теоретически разработанными являются статистические интервалы трёх видов: доверительные, толерантные и предикционные. Выбор вида интервала основывается на том, что представляет интерес: описание совокупности или процесса, из которого была отобрана выборка на контроль, или предсказание результатов контроля будущей выборки, отбираемой из той же самой совокупности или процесса, из которого была отобрана текущая выборка. К интервалам, используемым для описания особенностей распределения, относятся доверительные интервалы для параметров распределения (например, интервалы для математического ожидания и дисперсии нормального закона) и толерантные интервалы, содержащие, по крайней мере, заданную долю распределения. В противоположность им предикционные интервалы позволяют судить с заданным уровнем доверия в

каких пределах будет находиться, например, значение выборочного среднего или выборочной дисперсии в будущей выборке, отобранной из того же самого распределения.

Для целей контроля качества наиболее широко используются доверительные и, в несколько меньшей степени, толерантные интервалы. Самый наглядный пример – это процедуры приёмочного контроля качества по количественному признаку на основе нормативного уровня несоответствий, где доверительные интервалы для уровня несоответствий и толерантные интервалы для распределения значений показателя качества используются в правилах принятия решения по результатам контроля [1]. В то же время предикционные интервалы используются редко, хотя для конечного потребителя конкретной единицы продукции такие интервалы могут представлять больший интерес, чем, например, доверительные интервалы.

Ситуации, в которых предикционные интервалы для целей контроля качества всё-таки используются, относятся к случаю нормального или, по крайней мере, непрерывного распределения контролируемой характеристики [2]. Это означает, что соответствующие процедуры предполагают измерения контролируемой характеристики по какой-либо непрерывной шкале, то есть связаны с проведением контроля по количественному признаку. Однако, на практике широко распространены и процедуры контроля по альтернативному признаку, связанные с подсчётом числа несоответствующих единиц или числа несоответствий в выборке. В этом случае предикционные интервалы могут представлять не меньший интерес, но информации об особенностях их применения к результатам контроля по альтернативному признаку крайне мало.

Во-первых, необходимо учитывать, что при контроле по альтернативному признаку возможны два типа процедур контроля: 1) тип А – при контроле отдельных (изолированных) партий продукции; 2) тип В – при контроле последовательной серии партий, поступающих из непрерывного процесса производства с постоянным средним уровнем несоответствий. Очевидно, что условиям применения предикционных интервалов отвечает процедура типа В, при этом постоянство процесса производства должно подтверждаться с помощью контрольных карт. Во-вторых, результат контроля выборки X_n может быть выражен двумя способами: 1) в виде процента несоответствующих единиц продукции – в этом случае X_n представляет собой число несоответствующих единиц, обнаруженных при контроле выборки объёма n , и является биномиальной случайной величиной; 2) в виде числа несоответствий на 100 единиц продукции – в этом случае X_n рассматривается как число несоответствий, обнаруженных в выборе объёма n , и представляет собой случайную величину, описываемую распределением Пуассона.

Если $X_n = m$ – число несоответствующих единиц в выборке объёма n_1 , то границы $[\underline{Y}_n, \bar{Y}_n]$ $100(1 - \alpha)\%$ предикционного интервала для числа несоответствующих единиц Y_n в будущей выборке объёма n_2 могут быть найдены как наименьшее значение \underline{Y}_n и наибольшее значение \bar{Y}_n , удовлетворяющие условиям [3]:

$$1 - F_{Hy}(m - 1; n_1, m + \underline{Y}_n, n_1 + n_2) > \alpha / 2,$$

$$F_{Hy}(m; n_1, m + \bar{Y}_n, n_1 + n_2) > \alpha / 2,$$

где $F_{Hy}(x; n, M, N)$ – кумулятивная функция гипергеометрического распределения с параметрами n (объём выборки), M (число целевых элементов в совокупности) и N (объём совокупности). При этом, если $m = 0$, то $\underline{Y}_n = 0$, и если $m = n_1$, то $\bar{Y}_n = n_2$.

Если $X_n = m$ – число несоответствий в выборке объёма n_1 , то границы $[\underline{Y}_n, \bar{Y}_n]$ $100(1 - \alpha)\%$ предикционного интервала для числа несоответствий Y_n в будущей выборке объёма n_2 могут быть найдены как наименьшее значение \underline{Y}_n и наибольшее значение \bar{Y}_n , удовлетворяющие условиям:

$$1 - F_B(m - 1; m + \underline{Y}_n, n_1 / (n_1 + n_2)) > \alpha / 2,$$

$$F_B(m; m + \bar{Y}_n, n_1 / (n_1 + n_2)) > \alpha / 2,$$

где $F_B(x; n, p)$ – кумулятивная функция биномиального распределения с параметрами n (число испытаний) и p (вероятность «успеха» в отдельном испытании).

Для практического удобства применения приведённых зависимостей имеет смысл использовать соответствующее программное обеспечение, такое как язык программирования R [4]. В качестве примера на рисунке 1 приведён листинг функции для нахождения границ двустороннего предикционного интервала или односторонних предикционных пределов для числа несоответствующих единиц в будущей выборке.

```
PI_binom <- function(m, n1, n2, conf.level = 0.95,
                    type = c("interval", "lower", "upper")){

  if (length(type) > 1) {type <- "interval"}
  alpha <- if (type == "interval") {(1 - conf.level)/2} else {1 - conf.level}
  y <- 0:n2
  F1 <- phyper(m - 1, m + y, n1 + n2 - m - y, n1, lower.tail = FALSE)
  F2 <- phyper(m, m + y, n1 + n2 - m - y, n1)
  Y1 <- y[which.min(F1 < alpha)]
  Y2 <- y[which.min(F2 > alpha) - 1]
  if (m == 0) {Y1 <- 0}
  if (m == n1) {Y2 <- n2}
  switch(type,
         interval = {result <- c(Y1, Y2)},
         lower = {result <- Y1},
         upper = {result <- Y2})
  return(result)
}
```

Рис. 1. Возможный вариант функции на языке R нахождения границ предикционного интервала для числа несоответствующих единиц

Предположим, что на основе $X_n = 20$ дефектных интегральных микросхем, обнаруженных при контроле случайной выборки объёма $n_1 = 1000$, требуется найти границы 95% предикционного интервала для числа дефектных микросхем в будущей выборке объёма $n_2 = 1000$ из того же процесса производства. Результат применения функции для указанных условий приведён на рисунке 2, из

которого видно, что, как минимум с уровнем доверия 95% можно утверждать, что число дефектных микросхем в будущей выборке объёма $n_2 = 1000$ будет находиться в интервале [9, 35].

```
PI_binom(20, 1000, 1000)
```

```
## [1] 9 35
```

Рис. 2. Пример нахождения границ предикционного интервала

Приведённая функция также позволяет находить односторонние предикционные пределы – для этого необходимо задать соответствующее значение аргумента «type». Так, для рассматриваемого примера при type = "upper" будет получено значение $\bar{Y}_n = 32$. Таким образом, можно утверждать с уровнем доверия как минимум 95%, что число дефектных микросхем в будущей выборке объёма $n_2 = 1000$ не будет превосходить 32 единиц.

Аналогичная функция может быть написана и для нахождения границ двустороннего предикционного интервала или односторонних предикционных пределов для числа несоответствий в будущей выборке.

Список литературы

1. ГОСТ Р 50779.50 – 95. Статистические методы. Приёмочный контроль качества по количественному признаку. Общие требования. – М.: ИПК Издательство стандартов, 2004. – 21 с.
2. ГОСТ Р ИСО 16269-8 – 2005. Статистические методы. Статистическое представление данных. Определение предикционных интервалов. – М.: Стандартинформ, 2005. – 63 с.
3. Meeker W.Q., Hahn G.J., Escobar L.A. Statistical Intervals. A Guide for Practitioners and Researchers. – 2nd ed. – New Jersey: John Wiley & Sons, 2017.
4. R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>

Сведения об авторах:

Борбаць Николай Михайлович – к.т.н., доцент, доцент кафедры «Управление качеством, стандартизация и метрология»;

Школина Татьяна Викторовна – к.т.н., доцент, доцент кафедры «Информатика и программное обеспечение».