Решение задачи

Денисов Никита

18 февраля 2022 г.

Задача

Прибор для выявления брака на фабрике имеет вероятность ошибки 5% (и первого и второго рода), процент брака составляет 5% от всего объёма выпускаемой продукции.

1) Какая вероятность того, что мы выявили брак, если прибор выдал положительный результат - "продукция бракованная"?

2) Почему же в жизни все-таки используют такие приборы? Что можно было бы изменить в процедуре поиска брака, не меняя точности прибора, так, чтобы вероятность из первого вопроса P(брак|"+") выросла?

3) Какое правило можно сформулировать для выбора точности прибора, зная процент бракованной продукции?

$p = 0.05$ — вероятность брака.
$q_0 = 0.05$ — вероятность ложноположительного срабатывания прибора.
$q_1 = 0.05$ — вероятность ложноотрицательно срабатывания прибора.

1. Хотим посчитать $P(\text{Брак} \mid +)$. Воспользуемся формулой Байеса:
   $P(\text{Брак} \mid +) = \frac{P(+ \mid \text{Брак}) \cdot P(\text{Брак})}{P(+)}$. $P(+)$ посчитаем по формуле полной вероятности:

   $P(+) = p(1 - q_1) + (1 - p)q_0 = 0.05 \cdot 0.95 + 0.95 \cdot 0.05$

   $P(+ \mid \text{Брак}) = 1 - q_1 = 0.95$

   Итого, получаем:
   $P(\text{Брак} \mid +) = \frac{P(+ \mid \text{Брак}) \cdot p}{P(+)} = \frac{0.95 \cdot 0.05}{2 \cdot 0.95 \cdot 0.05} = \frac{1}{2}$

2. Такие приборы в жизни используют, чтобы отметать хотя бы какую-то часть бракованных изеделий. Но отвергаются также и изделия без брака. Это все равно лучше, чем допускать все изделия, ведь брак может привести к печальным последствиям.
   Если точность прибора изменить нельзя, то можно попробовать запускать прибор несколько раз на каждой продукции. При этом есть вариации как именно запускать прибор. Если мы хотим, например, допустить минимальное количество брака, то лучше запускать прибор не более k раз до первого положительного срабатывания и тогда отметать, а если все k были отрицательные то оставлять. Если же наша цель более точно определять бракованная ли деталь, то стоит k раз запускать прибор вне зависимости от предыдущих результатов и в зависимости от количества положительных срабатываний определять бракованная ли деталь.
   Пусть мы сделали k запусков на одной продукции и получили q положительных результатов. Определим положительный результат прибора, если $q \geq t(k)$. Тогда посмотрим:

$P(\text{Брак} \mid +) = \frac{P(+ \mid \text{Брак}) \cdot p}{P(+)}$
$P(+) = p \cdot \sum_{q=t(k)}^{k} \binom{k}{q} (1-q_1)^q \cdot q_1^{k-q} + (1-p) \cdot \sum_{q=t(k)}^{k} \binom{k}{q} q_0^q (1-q_0)^{k-q}$ — формула полной вероятности + перебрали количество положительных срабатываний, которые нас устраивают.

$P(+ \mid \text{Брак}) = \sum_{q=t(k)}^{k} \binom{k}{q} (1-q_1)^q \cdot q_1^{k-q}$

Чтобы $P(\text{Брак} \mid +) > \frac{1}{2}$ нужно чтобы выполнялось:
$A = p \cdot \sum_{q=t(k)}^{k} \binom{k}{q} (1-q_1)^q \cdot q_1^{k-q} > (1-p) \cdot \sum_{q=t(k)}^{k} \binom{k}{q} q_0^q (1-q_0)^{k-q} = B$

Напишем скрипт на Python и проверим какие $k, t(k)$ нам подходят. (Перебрал k до 100 и t вплоть до k максимизируя величину $A - B$)

Скрипт выдал результат, что при $k = 45, t(k) = 23$ величина

$(A - B)$ оказывается близка к 0.05 (0.04999999999999991).
$P(\text{Брак} \mid +) = \frac{A}{A+B} \sim \frac{B+0.05}{2B+0.05} > \frac{B+0.05}{2B+0.1} = \frac{1}{2}$.

Естественно, данные значения не единственные подоходящие и проверять одну продукцию 45 раз может быть слишком затратно. В таком случае нужно искать баланс между количеством проверок и получающейся вероятностью.

3. Теперь зная p определим какие q_0, q_1 нужны для прибора чтобы получить $P(\text{Брак} \mid +) \geq x$. Мы уже выяснили:
   $P(+) = p(1-q_1) + (1-p)q_0$
   $P(\text{Брак} \mid +) = \frac{P(+ \mid \text{Брак}) \cdot p}{P(+)} = \frac{p(1-q_1)}{p(1-q_1)+(1-p)q_0} \geq x$. Выразим q_0 и определим для него границу при фиксированном q_1:
   $\frac{p(1-q_1)(1-x)}{(1-p)x} \geq q_0$
   Можем и наоборот, выразить q_1 и определить границу для него при фиксированном q_1:
   $p - pq_1 \geq -xpq_1 + px + xq_0 - xpq_0 \Leftrightarrow -q_1 p(1-x) \geq p(x-1) + q_0 x(1-p) \Leftrightarrow -q_1 \geq \frac{p(x-1)+q_0 x(1-p)}{p(1-x)} \Leftrightarrow q_1 \leq \frac{p(1-x)-q_0 x(1-p)}{p(1-x)}$

   Зная эти границы можно экспериментальным путем определить подходящие q_0, q_1 по заданному x, например, написав скрипт.