

Reverse cocaine sting

In 1981 in Florida, police raided a house and found 496 bags containing white powder. They tested four random bags and the results were positive for cocaine. The bags were then either destroyed or taken into evidence, depending on which account you read.

Two of the remaining 492 bags were used in a sting operation and were sold to an alleged dealer. Let's say his name was Tony. The bags were sold, but before Tony was taken into custody, the two sold bags went missing. It is not clear what happened to the other 490 bags. Presumably they were destroyed as well.

The two bags that were sold to Tony were never tested and never recovered, so what is the probability that Tony actually bought some legal substance that looks like cocaine, e.g. baking soda?

Let X represent the number of bags containing cocaine that were seized in the original raid and let A be the event that $X = 4$. Let Y represent the number of bags containing a legal substance that were purchased by Tony and let B be the events that $Y = 2$. The probability that Tony actually purchased as legal substance is

$$P(A \cap B) = P(A)P(B|A) = P(X = 4)P(Y = 2|X = 4)$$

The probabilities $P(A)$ and $P(B|A)$ depend on how many bags out of the original 496 actually contained cocaine. At least 4 did. If exactly 4 did, then $P(B|A) = 1$, but $P(A)$ is very close to 0, so $P(A \cap B)$ is also close to 0.

Let K be the number of bags out of the original 496 that contained cocaine. Then $4 \leq K \leq 496$. As K increases, $P(A)$ increases, but $P(B|A)$ decreases. When $K = 496$, $P(B|A) = 0$ and $P(A \cap B) = 0$. What is the value of K for which $P(A \cap B)$ is maximal, and what is $P(A \cap B)$, the probability of “innocence?”

Hypergeometric distributions

Draw n times from a population of size N without replacement. If the population contains K successes, then the probability of obtaining k successes in the n draws is

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

In the special case where $n = k$, the formula reduces to

$$P(X = k) = \frac{\binom{K}{k}}{\binom{N}{k}}$$

Reverse cocaine sting solution

$$P(X = 4) = \frac{\binom{K}{4}}{\binom{496}{4}}$$
$$P(Y = 2|X = 4) = \frac{\binom{496-K}{2}}{\binom{492}{2}}$$
$$P(Y = 2 \cap X = 4) = P(X = 4) \cdot P(Y = 2|X = 4) = \frac{\binom{K}{4} \binom{496-K}{2}}{\binom{496}{4} \binom{492}{2}}$$

R Code

There is an R function, `dhyper`, that gives the values from a hypergeometric distribution. The signature is `dhyper(k, K, N - K, n)` where N is the number of successes in the population, N is the population size, n is the number of trials, and k is the number of successes in n trials. Thus,

$$P(X = 4) = \text{dhyper}(4, K, 496 - K, 4)$$

and

$$P(Y = 2|X = 4) = \text{dhyper}(2, 496 - K, K - 4, 2)$$

There is a Jupyter notebook containing the R code at https://github.com/AaronGladish/stats_rco. If you have Jupyter notebooks on your desktop, you can download the .ipynb file and run it.