Jak počítá statistik ryby v rybníku

V rybníku je neznámý počet ryb (označme ho N), o němž bychom se rádi něco dozvěděli. Metoda přímého, vyčerpávajícího měření (populační šetření), vyžadující, aby byl rybník vypuštěn, není přitom často použitelná.

Statistik navrhuje nepřímý postup, opírající se náhodný o výběr:

Z rybníka se vyloví M ryb, které se označí a pustí do rybníka zpět. Po nějaké době (zaručující promíchání M označených ryb mezi ostatní) se provede další částečný výlov; řekněme, že při něm bylo vyloveno n ryb, mezi nimiž se rozpoznalo k $(0 \leq k \leq \min(M; n))$ ryb dříve označených. Přijme-li se princip, že podíl označených ve vzorku n vylovených ryb se rovná podílu označených v celém rybníku, měl by platit vztah

$$\frac{k}{n} = \frac{M}{N}$$

v němž k, M a n jsou veličiny známé, což umožňuje odhadovat počet ryb v rybníku ze vzorce po jednoduché úpravě:

$$N = \frac{M \cdot n}{k}.$$

Profesionální statistický přístup tímto výpočtem ovšem nekončí – spíše teprve začíná. Získaný výsledek N je třeba považovat za odhad, který je vystaven riziku omylu. Bylo by přece naivně optimistické předpokládat, že v rybníku žije právě tolik ryb, kolik ukazuje náš odhad. Podstatné je, že velikost chyby (odchylky odhadu od neznámé skutečnosti) se dá rovněž statistickými metodami posoudit a ukázat její závislost na rozsahu výběru n a podílu označených ryb M/N.

Z toho vyplývá, že počet k označených ryb ve druhém výlovu je náhodná veličina K s hypergeometrickým rozdělením $K \sim \mathsf{H}(M; N; n)$ pro $k \in (\max(0; n - (N - M)); \min(M; n))$, kde pravděpodobnost, že náhodná veličina K nabyde hodnoty k spočteme jako

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

a střední hodnotou jako $\mathsf{E}K = \frac{M \cdot n}{N}$, což naznačuje, že výše uvedený postup odhadování byl rozumný.

Volně podle http://medinfo.upol.cz/PUVODNI/statistika/