T711 (Parlar)

# Calculation of the "correct selection" probabilities in the Coke-Pepsi experiment

Let's denote the choose(n,r) function in R by $\binom{n}{r}$, the binomial formula. So, $\binom{4}{2}=6$, etc.

| | C | | P |
|---|---|---|---|
| **0 correct** | PPPP | Only one way | CCCC |
| **1 correct in each block** / 2 correct total | $\underbrace{\text{CPPP}}$ This can happen in $\binom{4}{1}=4$ ways: PCPP, etc | | PCCC This can happen in $\binom{4}{1}=4$ ways: CPCC, etc. |

But, this results in $4\times4=16$ possibilities: (CPPP, PCCC), (PCPP, CPCC), etc.

| | CPPP | PCPP | PPCP | PPPC |
|---|---|---|---|---|
| CPPP | · | · | · | · |
| PCPP | · | · | · | · |
| PPCP | · | · | · | · |
| PPPC | · | · | · | · |

16 in total

2 correct in each block

4 correct total

C

P

$\underbrace{CCPP}$ Can happen in $\binom{4}{2} = 6$ ways

$\underbrace{PPCC}$ Can happen in $\binom{4}{2} = 6$ ways

This results in 6x6 possibilities (CCPP, PPCC), (CPCP, PCPC), etc.

| | CCPP | CPCP | CPPC | PCCP | PCPC | PPCC |
|---|---|---|---|---|---|---|
| CCPP | . | . | . | . | . | . |
| CPCP | . | . | . | . | . | . |
| CPPC | . | . | . | . | . | . |
| PCCP | . | . | . | . | . | . |
| PCPC | . | . | . | . | . | . |
| PPCC | . | . | . | . | . | . |

3 correct in each block

6 correct total

C

P

This is the same as 1 correct in each block. Happens in $\binom{4}{3} = 4$ ways

Total of $4 \times 4 = 16$ possibilities

4 correct in each block

8 correct total

C

CCCC

P

PPPP

Only one way

Thus,

| # Correct | 0 | 2 | 4 | 6 | 8 |
|---|---|---|---|---|---|
| Prob | $\frac{1}{70}$ | $\frac{16}{70}$ | $\frac{36}{70}$ | $\frac{16}{70}$ | $\frac{1}{70}$ = p-value |

If $H_0$: Subject guessing and if he/she gets all correct, I would Reject $H_0$ since the prob. of this outcome is extremely low ($\frac{1}{70}=1.4\%$)

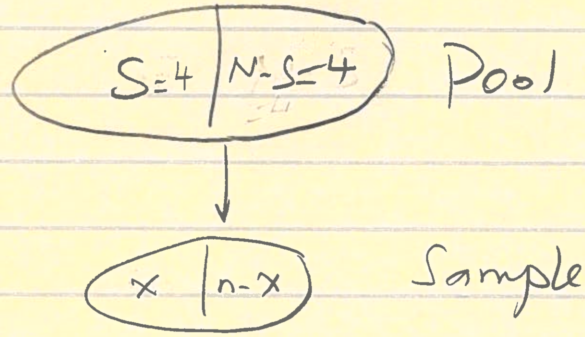
Note: There is an easy way to calculate these prob's once we see that we only need to focus on the # correct in a batch of four.

This involves using the hypergeometric distribution (which generalizes the binomial distribution).

Let's see →

We have a "pool" of $N=8$ cups. The pool contains $S=4$ Cokes (C) and $N-S=4$ Depsis (P). We draw a random sample of (first) $n=4$ from the pool.

Define $X$ as the # of successes (correct identifications). Clearly $X$ can be 0, 1, 2, 3 or 4.




It can be shown that (in more advanced courses),

$$\Pr(X=x) = \frac{\binom{S}{x}\binom{N-S}{n-x}}{\binom{N}{n}}$$

where $\binom{a}{b} = \frac{a!}{b!\,(a-b)!} = \text{choose}(a,b)$

and $a!$ is the factorial.

Rcmdr does this nicely, as follows

Distributions > Discrete distributions
> Hypergeometric distribution
> " probabilities

| | | | | |
|---|---|---|---|---|
| m (...) | $\boxed{4}$ | Our C | S | notation here |
| n (...) | $\boxed{4}$ | " P | N-S | |
| k (...) | $\boxed{4}$ | Sample | n | |

| | Prob | | |
|---|---|---|---|
| ⇒ 0 | .014 | 0 success in 4 | ⇒ 0 in 8 |
| 1 | .228 | 1 " | 2 " 8 |
| 2 | .514 | 2 " | 4 " 8 |
| 3 | .228 | 3 " | 6 " 8 |
| 4 | .014 | 4 " | 8 " 8 |