Project Ideas

Elliott Pryor

10/19/2020

Hyper Geometric

1.

Given that we tagged r turtles. We want to estimate the population size. The workers re-capture n turtles randomly. We count which ones have already been tagged. We determine that the number of already tagged turtles is s. What population size N would give us this expected number of turtles?

Note we could plug in numbers for r, s, n

We are looking for $E(Y) = s = \frac{nr}{N}$.

$$N = \frac{nr}{s}$$

2.

If some wizard tells us that the population is actually N', what is the probability that we recapture s turtles?

$$P(Y = s) = \frac{\binom{r}{s} * \binom{N'-r}{n-s}}{\binom{N'}{n}}$$

3.

Suppose sally has x favourite marked turtles. Suppose the populatin size is N'. How many turtles does she need to catch to have a 50% chance of catching one of her turtles?

So this is really simple (but wording may make it confusing). You don't have to consider the original r marked turtles. Instead just the x marked turtles of Sally's. So $Y \sim Hypergeometric(N', x, n)$.

We can probably compute this by hand, but also can use the r function:

```
# placeholder values so not error
N = 1000 # N' variable (pop size)
x = 3 # sallys favorites
n = 100 # sample size

qhyper(p = 0.5, m = x, n = N - x, k = n)
```

```
## [1] 0
```

4.

If there are N' turtles and we want to catch n of them at a time. How many have to be marked so that we expect to catch a turles each time.

$$E(Y) = a = n\frac{r}{N} \implies r = \frac{aN}{n}$$

Normal

I have no idea how to relate this to turtles so just writing generally

1.

Suppose that data is distributed normally. We know that exactly 95% of the data has a value between $[a, b]$. What is the mean and variance?

$\mu = \frac{a+b}{2}$

$\sigma^2 = \left(\frac{b-\mu}{2}\right)^2$

2.

Find the peak value of the pdf of normal distribution.

We know it occurs at $y = \mu$ so the peak value is

$\frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-1}{2\sigma^2}(0)^2\right) = \frac{1}{\sqrt{2\pi\sigma}}$

Not the best question, but IDK what you can even ask with normal that is interesting

3.

Given normal distribution centered at μ. We know that exactly 68% of the data falls in $[a, b]$. What range of values can we expect 98% of data to fall between.

$\sigma = b - \mu$, so we expect that 98% of data to be in $[a - \sigma, b + \sigma]$

4.

What is the probability that a value is $v > \mu + \sigma$ on a normal distribution centered at μ with standard deviation σ.

I like this one

We can use the empirical rule even though it is 1 sided. Since it is symmetric, we know by the empirical rule that $1 - 0.68 = 0.32$ of the data fall outside of 1 standard deviation. Since it is symmetric, half of this is on the left hand side. So we only care about the portion that is on the right hand side. So $0.32/2 = 0.16$