Up until now → asymptotic inference 2/8/22

* Requires large samples
→ CLT → approximate CIs & p-values using normal approximation

R: prop.test

- One sample & 1 binary variable → exact binomial inference (p-value)
  R: binom.test

Exact inference for 2 proportions (2 binary variables):

1. Fisher's Exact Test
2. Randomization (simulation-based) tests * Not in book

Fisher's Exact Test

Example: Penguins -

Scope of Inference → ① Cause-&-effect?

Randomized experiment → $\boxed{\text{Yes}}$

Observational Study → No

② To whom can we generalize?
Representative sample of what population?

- Only generalize to penguins similar to those in the sample

Data:

	Survive	Not	
Metal	$\boxed{3}$	7	10
Not metal	6	4	10
	9	11	20

row totals fixed → "binomial" sampling

Consider count in 1^{st} cell: $n_{11} = 3$

H_0: No association between survival status & whether the penguin had a metal band.

$\pi_1 = P(\text{survive} \mid \text{metal band})$

$\pi_2 = P(\text{survive} \mid \text{no metal band})$

$H_0: \pi_1 - \pi_2 = 0$

$H_0: \pi_1 / \pi_2 = 1$

$$\boxed{H_0: \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} = 1}$$

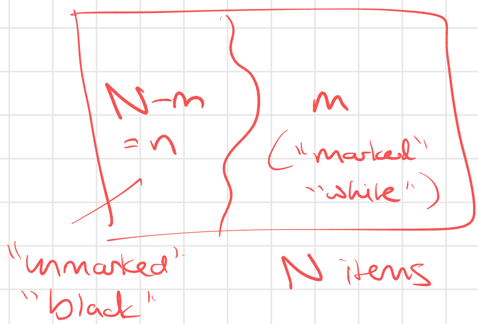
$$\text{Odds ratio} = \theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$$

$H_0: \theta = 1$

$H_a: \theta < 1$

Under H_0 $\Rightarrow$ Model n_{11} with a hypergeometric distribution!

Scenario: Finite population of N items.



Parameters:

m = # marked in pop.

n = # unmarked in pop.

k = sample size

- Sample k items without replacement.

X = # of marked items in the sample

Note: Sampling with replacement.

$$X \sim \text{Bin}\left(k, \frac{m}{n+m}\right)$$

	Success	Failure	
Group 1	n_{11}	n_{12}	n_{1+}
Group 2	n_{21}	n_{22}	n_{2+}
	n_{+1}	n_{+2}	n

Sample n_{1+} from population of n items

n_{+1} marked
n_{+2} unmarked

Assuming H_0: No association

X = # of marked items that end up in our sample.

$$P(X = x) = \frac{\text{\# of samples where } x \text{ marked \& } k-x \text{ unmarked}}{\text{total \# of possible samples}}$$

$$= \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}} = \frac{\binom{n_{+1}}{n_{11}}\binom{n_{+2}}{n_{1+}-n_{11}}}{\binom{n}{n_{1+}}}$$

R: dhyper(x, m, n, k)

dhyper(n_{11}, n_{+1}, n_{+2}, n_{1+})

p-value $= P(X \leq 3)$

R: phyper(3, 9, 11, 10)

row 1 total

col. 1 total

col. 2 total

Randomization Test — Simulate the random assignment process 1000's of times assuming H_0.

p-value = proportion of those simulations which resulted in the observed data or something more extreme

Ex:

	Survive	Not	
Metal	3	7	10
No metal	6	4	10
	9	11	20

Under H_0 ↳ 9 penguins would survive regardless of group assignment

Simulate: 9 survivors, 11 non-survivors → 10, 10

R.A.

How to simulate w/ cards?

Total # of cards? 20 — 1 for each penguin

blue (survivors) 9

red (non-survivors) 11

Shuffle

10 metal

10 non-metal