

Exercise set 10.3: Organizing data

- A survey of 110 lung cancer patients showed that 70 were cigarette smokers, 60 lived in urban areas, and 35 had hazardous occupations. Forty of the smokers lived in urban areas, 15 had hazardous occupations, and 5 were in both categories. Ten of the patients with hazardous occupations neither lived in an urban area, nor smoked.

- How many of the patients living in urban areas had hazardous occupations?
- How many of those living in the urban areas neither smoked nor had hazardous occupations?
- How many patients neither smoked, nor lived in an urban area, nor had a hazardous occupation?

See the precast and PDF in the week 11 resources for solution.

- A survey was conducted of tree species in three forest types in the Katarniaghat tree sanctuary in Uttar Pradesh, India. The forest types and number of species observed in each were: Teak plantation 25, Sal forest 30, and Miscellaneous forest 40. 18 species were common to the Miscellaneous and Sal forest, 15 were common to the Teak and Sal forests, and 18 were common to the Teak and Miscellaneous forests. 4 species were found only in the Teak and Sal forests.

- What is the probability that a given species would be found in each forest type?

$$P(T) = 25/55 = 5/11 \quad P(S) = 30/55 = 6/11 \quad P(M) = 40/55 = 8/11$$

- What is the probability that a given species would be found only in the Teak plantation?

$$P(T \cap \bar{S} \cap \bar{M}) = 3/55$$

- What is the probability that a given species would be found in more than one forest type?

$$P = (4 + 11 + 7 + 7)/55 = 29/55$$

- What is the probability that a given species would be found in only one forest type?

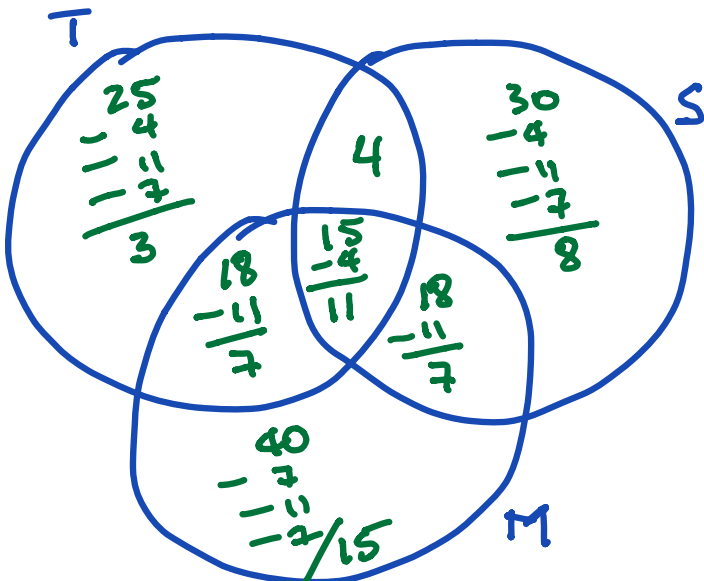
$$P = (3 + 8 + 15)/55 = 26/55$$

- What is the probability that a given species would be found in all the forest types?

$$P(T \cap S \cap M) = 11/55$$

⑧ Total

$$\begin{array}{r} 3 \\ 4 \\ 8 \\ 7 \\ 11 \\ 7 \\ 15 \\ \hline 55 \end{array}$$



$$\begin{array}{ll} \cancel{n(T) = 25} & \textcircled{5} \\ \cancel{n(S) = 30} & \textcircled{6} \\ \cancel{n(M) = 40} & \textcircled{7} \\ \cancel{n(M \cap S) = 18} & \textcircled{3} \\ \cancel{n(T \cap S) = 15} & \textcircled{2} \\ \cancel{n(T \cap M) = 18} & \textcircled{4} \\ \cancel{n(T \cap S \cap M) = 4} & \textcircled{1} \end{array}$$

order of use ↑

3. Since brain diseases all attack the same organ, perhaps they share genetic similarities that enable their attacks. Researchers identified 4,216 genes associated with the three most common brain diseases: Alzheimer's, Multiple Sclerosis (MS), and Strokes. 464 were associated with Alzheimer's, 1,101 with MS, of which 50 were common to both. 708 genes were unique to MS and 318 were unique to Alzheimer's. Alzheimer's and Strokes had 110 genes in common, while 357 were common to Strokes and MS. Finally just 4 genes were common to all three diseases.

- a. What is the probability that a gene of one of the diseases is shared with another disease?

$$P = (46 + 4 + 106 + 353) / 4216 = 509 / 4216 \approx 0.12$$

- b. What is the probability that a gene is uniquely associated with a single disease?

$$P = (318 + 708 + 2681) / 4216 = 3707 / 4216 \approx 0.88$$

- c. What is the probability that a gene is associated with MS?

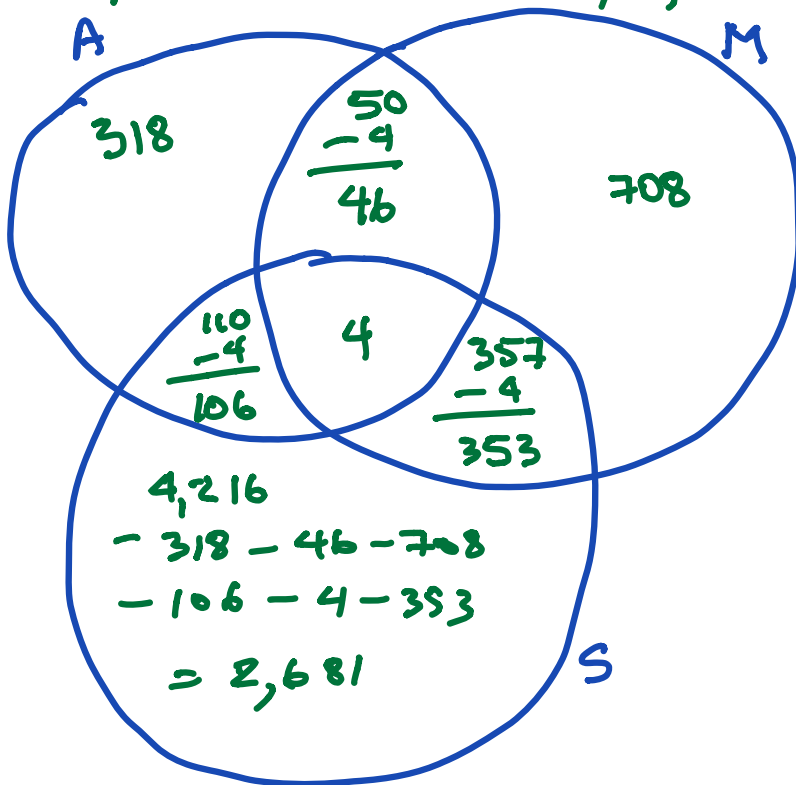
$$P = 1,101 / 4,216 \approx 0.26$$

- d. What is the probability that a gene associated with Strokes is also associated with MS?

$$P(M/S) = (4 + 353) / (106 + 4 + 353 + 2,681) \approx 0.11$$

- e. What is the probability that a gene is not associated with Strokes?

$$P(\bar{S}) = (318 + 46 + 708) / 4216 \approx 0.25$$



$$n(\omega) = 4,216 \quad (2)$$

$$\textcircled{8} n(A) = 464 \quad \times$$

$$\textcircled{9} n(M) = 1,101 \quad \times$$

$$\cancel{n(A \cap M) = 50} \quad (4)$$

$$\cancel{n(M \cap \bar{A} \cap \bar{S}) = 708} \quad (1)$$

$$\cancel{n(A \cap \bar{M} \cap \bar{S}) = 318} \quad (2)$$

$$\cancel{n(A \cap S) = 110} \quad (5)$$

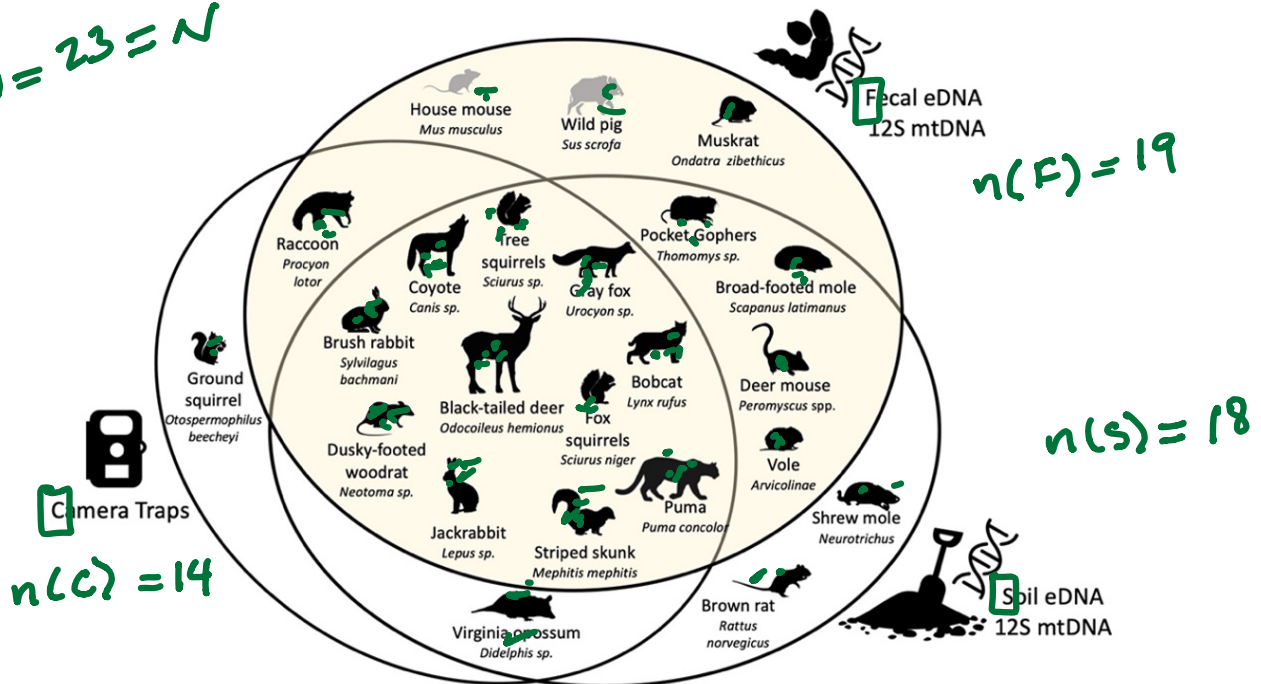
$$\cancel{n(M \cap S) = 357} \quad (6)$$

$$\cancel{n(A \cap M \cap S) = 4} \quad (3)$$

check: Does $318 + 46 + 4 + 106 = 474 \neq 464$ No!

Does $708 + 46 + 4 + 353 = 1,111 \neq 1,101$ No!

- $$n(u) = 23 = N$$



- Fecal DNA. $p(F) = 19/23 \approx 0.83$

- $$P(C \cap F \cap S) = 11/23 \approx 0.48$$

- $$p(c) = 14/23 \approx 0.61$$

- $$p = (1 + 2 + 3) / 23 \approx 0.26$$

- probability? *Oop! Duplicate. See a.*

- photographed? $p(c/s) = 12/18 = 2/3 \approx 0.67$

- Unknown! No data provided.

- What is the probability that a bobcat was detected by all three methods?
1.0! Because one was so no uncertainty.

- o.o! " " wasn't " " "