Comparing Samples

Logistics

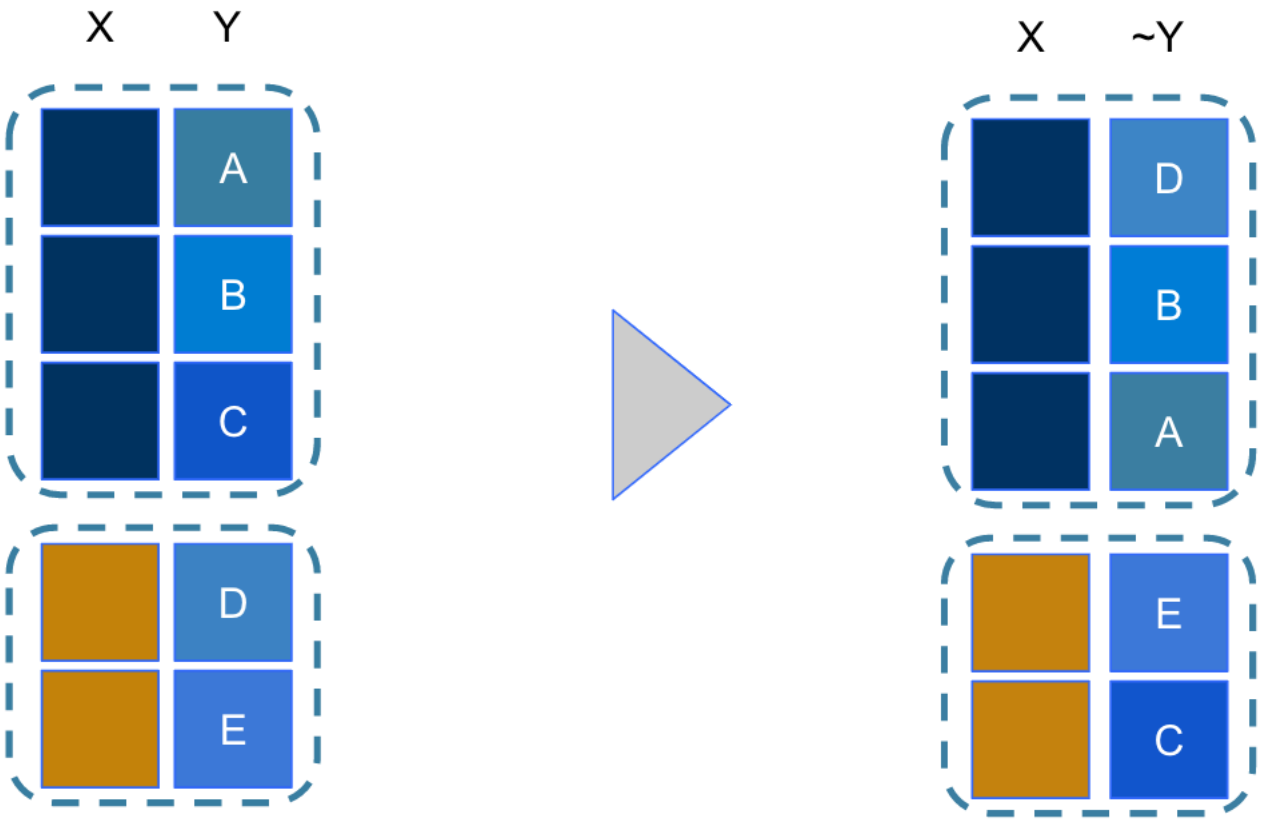
- Assignment out today, due next Thursday

Deflategate

Tom Brady on Deflategate: 'I've just moved on, man'

Adam Kurkjian Sunday, October 09, 2016

Permutation



Permutation Test

- Whether two samples are drawn randomly from the same underlying distribution

- If the null is true, all rearrangements of the attribute values among the two classes are equally likely

- So compute the observed test statistic
  - Then shuffle the attribute values and recompute statistic
  - Repeat
  - Compare

Official Analysis

- “The average pressure drop of the Patriots game balls exceeded the average pressure drop of the Colts balls by 0.45 to 1.02 psi, depending on various possible assumptions regarding the gauges used, and assuming an initial pressure of 12.5 psi for the patriots balls and 13.0 for the Colts balls.”

A / B Testing

- Two random samples:
  - Sample A
  - Sample B
  - E.g., Patriots vs Colts footballs

- Question: are they drawn from the same underlying distribution?

A / B Tesitng

- Null: The two samples are drawn from the same underlying population distribution; they look like two random draws from the same set

- Alternative: The samples are drawn from different distributions; they don’t look like random draws from the same set.

- Approach: permutation test (all rearrangements of the variable values among the two samples are equally likely, so shuffle, compute test statistic, repeat, compare)

The Test Statistic

- If the samples are categorical, then a natural test statistic is the total variation distance – it measures the distance between the distributions in two samples

- If the samples are numerical, often a simpler statistic, such as the absolute difference between the two sample means, is fine

Activity – Maternal Smokers

- There's a csv in cs1070.com/cs1070_materials/demos/ called baby.csv

- Work in your groups to perform an analysis of whether there is a significant statistical difference in baby weight for people who smoke vs people who don't

To Turn In

- 1 page write up:
  - Your group member's names
  - Explain your analysis (what was your null hypothesis, your alternative hypothesis, and your test statistic)
  - Describe your sampling strategy
  - State your P-value and whether you reject the null hypothesis or not
  - Include relevant and motivating figures
- Print this + your notebook
- Turn in at the beginning of class next Tuesday
- Will count as a quiz grade