Nonparametric statistics

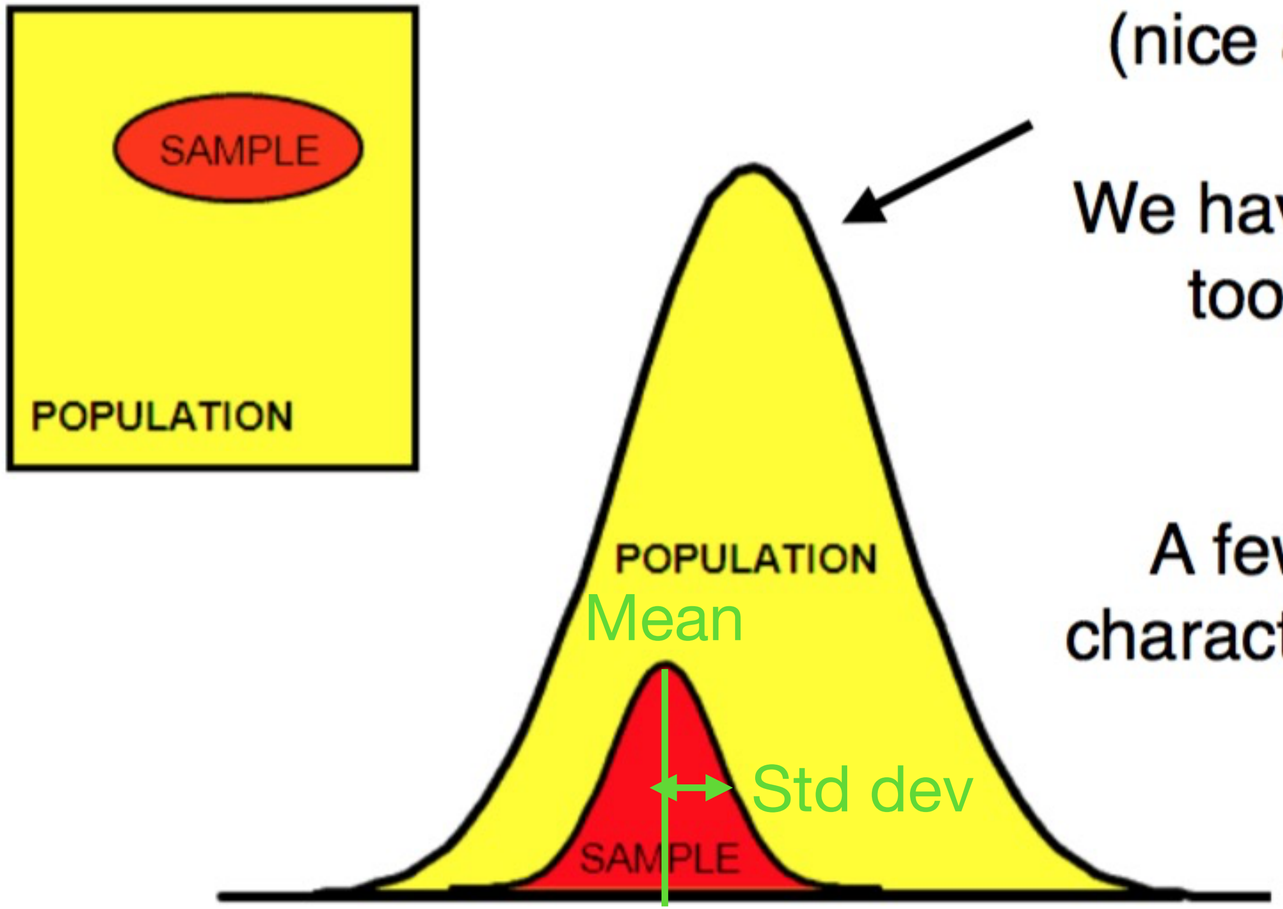
Jason G. Fleischer, Ph.D.
Asst. Teaching Professor
Department of Cognitive Science, UC San Diego

jfleischer@ucsd.edu

@jasongfleischer

https://jgfleischer.com

Non-parametric Statistics



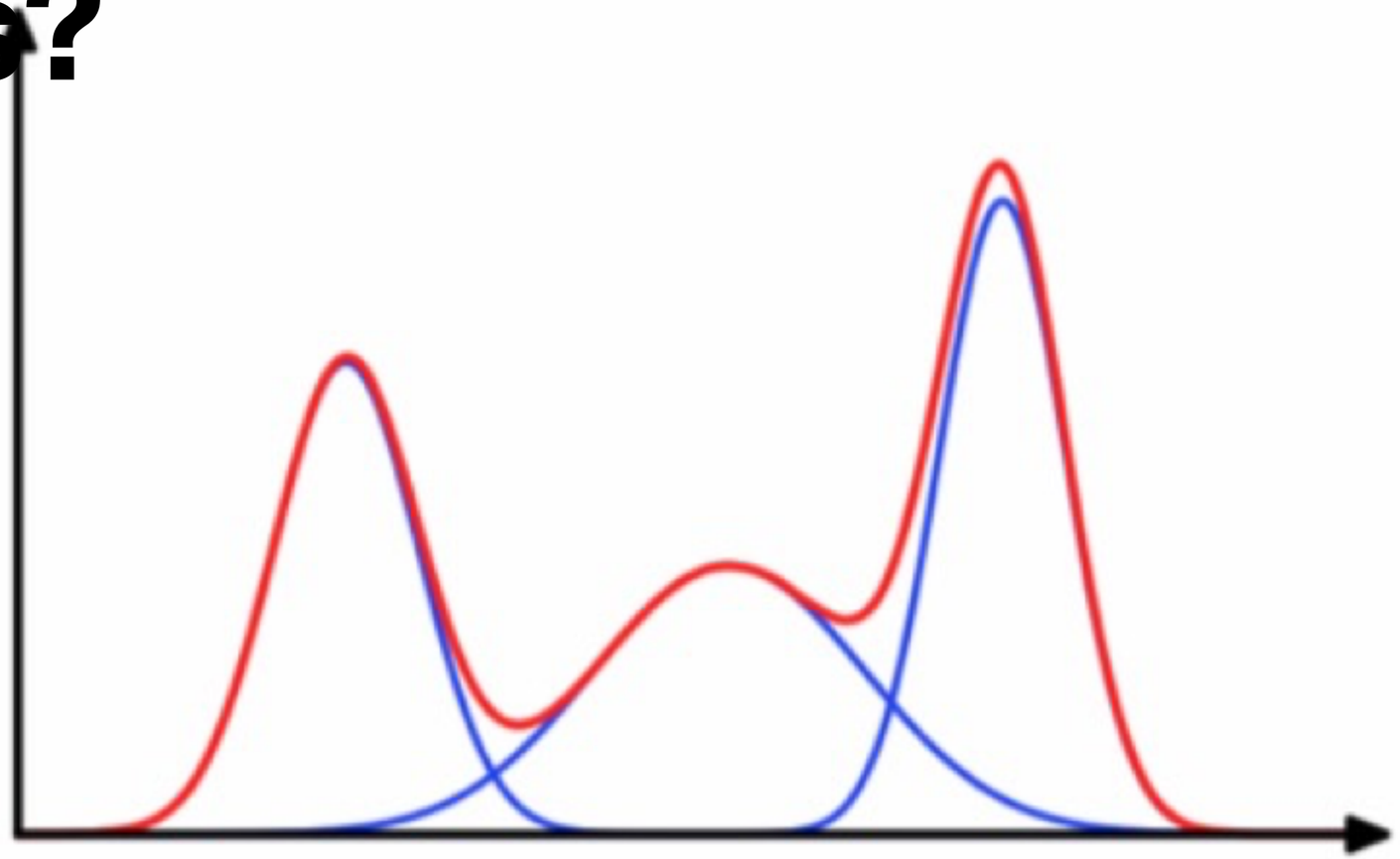
Normal distribution
(nice and friendly)

We have good math tools for this.

A few parameters **fully** characterize the distribution.

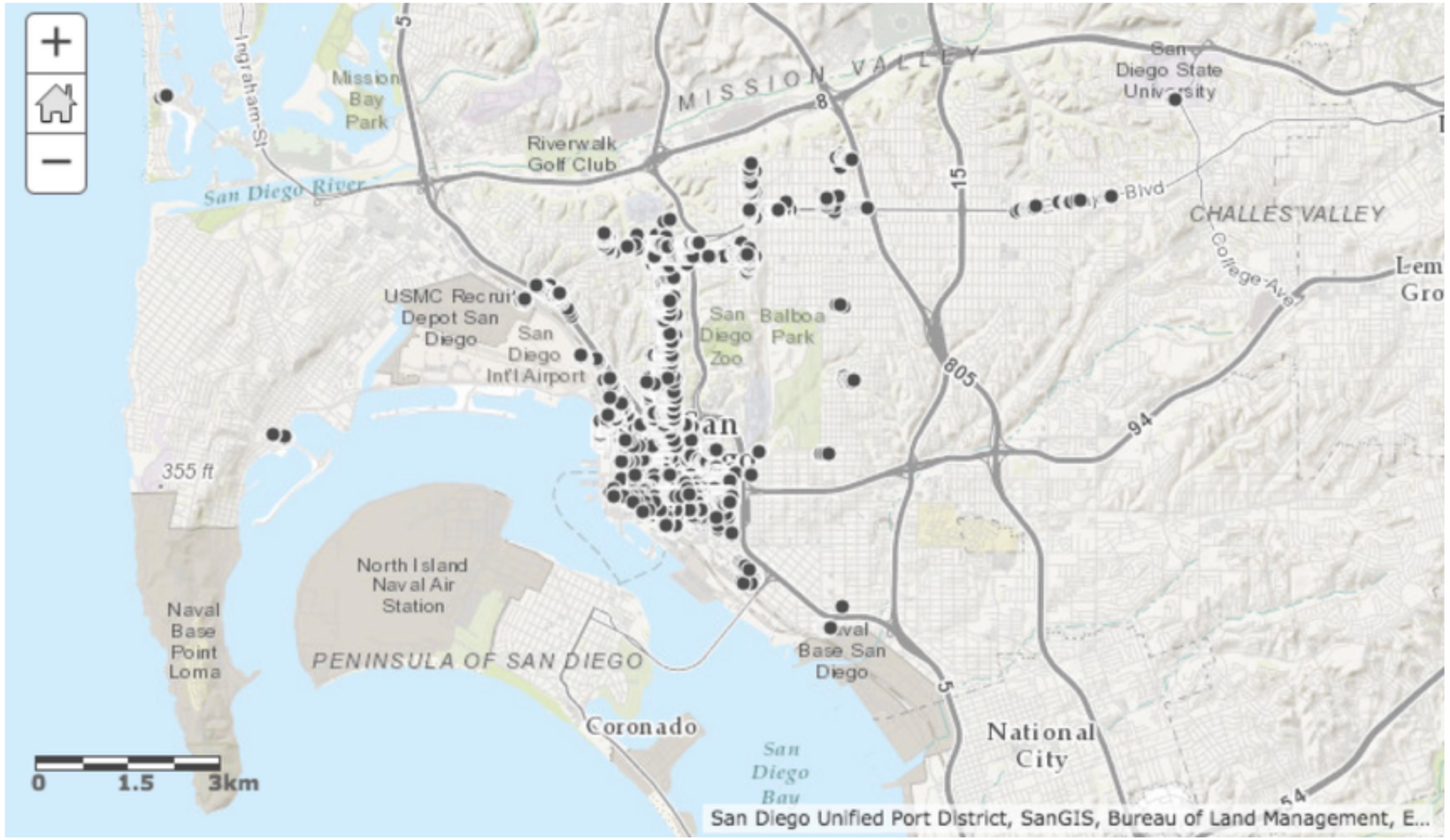
adapted from Brad Voytek

Non-parametric Statistics: What if your distribution looks like this?



adapted from Brad Voytek

Non-parametric Statistics: ...or like this?



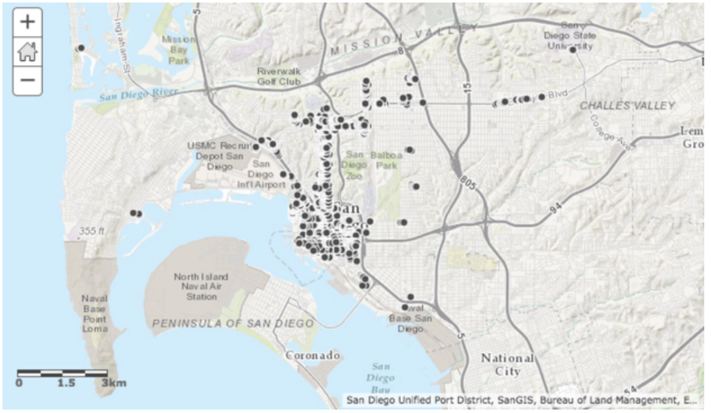
Parameters (like mean and variance) cannot fully and accurately capture this distribution!

Hence, we require **non-parametric statistics**.

adapted from Brad Voytek

When to turn to non-parametric statistics...

When underlying distributions are non-normal, skewed, or cannot be parameterized simply.



Like	Like Somewhat	Neutral	Dislike Somewhat	Dislike
1	2	3	4	5

Non-parametric Statistics: distribution-free

Myth: Non-parametric statistics does not use parameters.
Fact: Non-parametric statistics does not make *assumptions about* / parametrize the underlying distribution generating the data.

“Distribution-Free” statistics

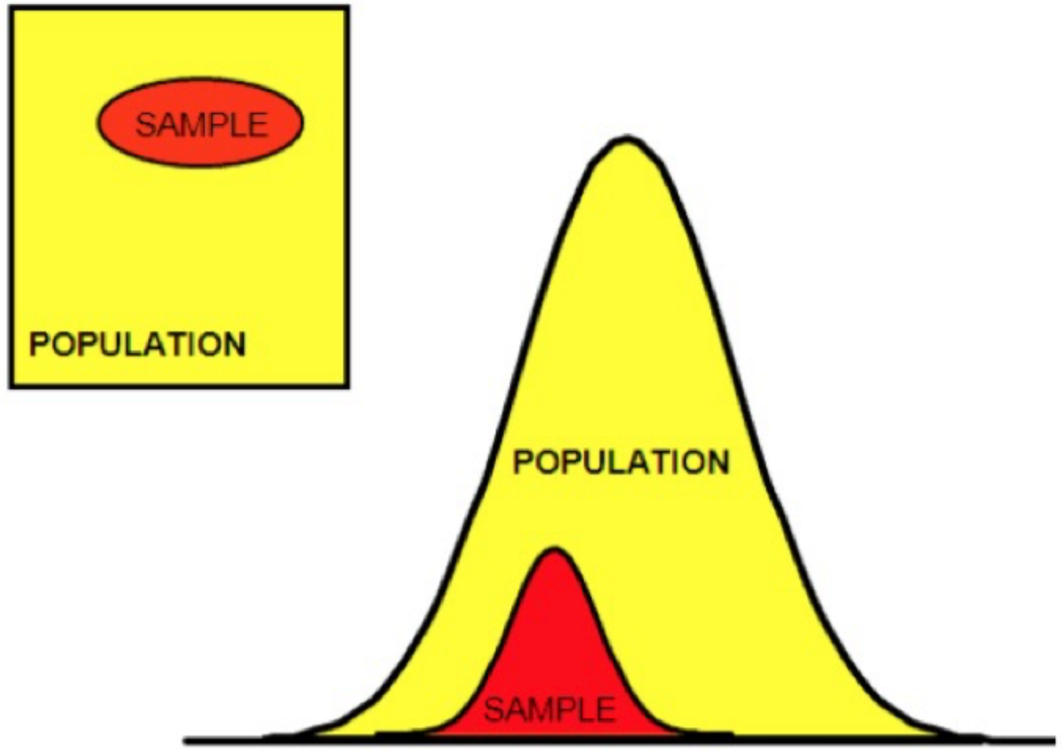
Meaning, it does not assume data-generating process (like heights) result in, *e.g.*, normally-distributed data

Resampling statistics: The What

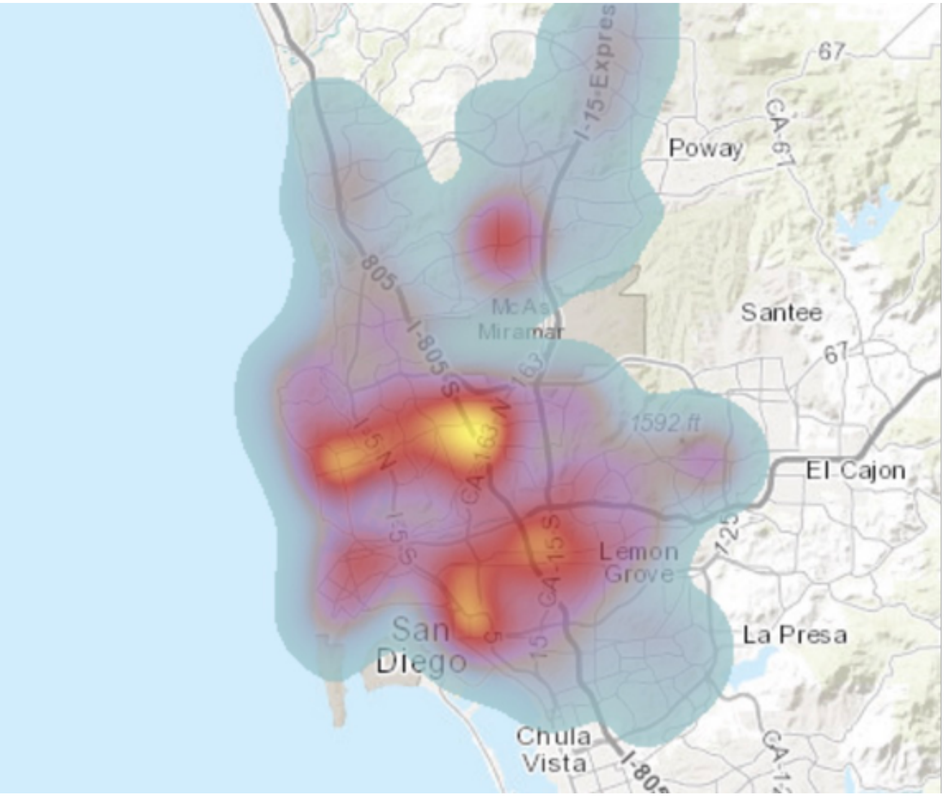
Empirical null distribution (Monte Carlo)
Rank Statistics (Mann Whitney U)
Kolmogorov-Smirnoff Test
Non-parametric prediction models

1) Monte Carlo

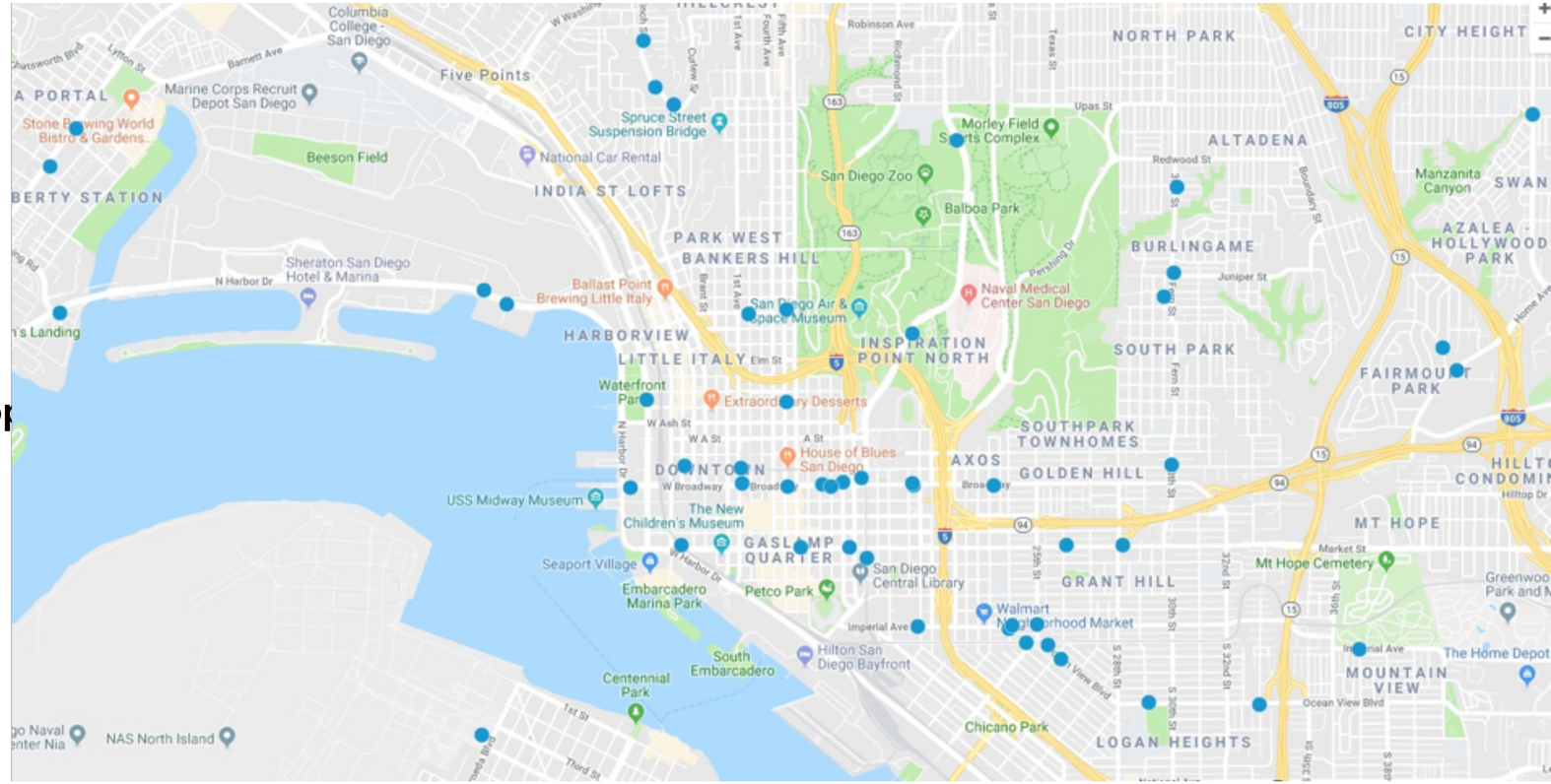
- How can we build a more realistic “null distribution” for the sample estimate without knowing the population it’s drawn from?



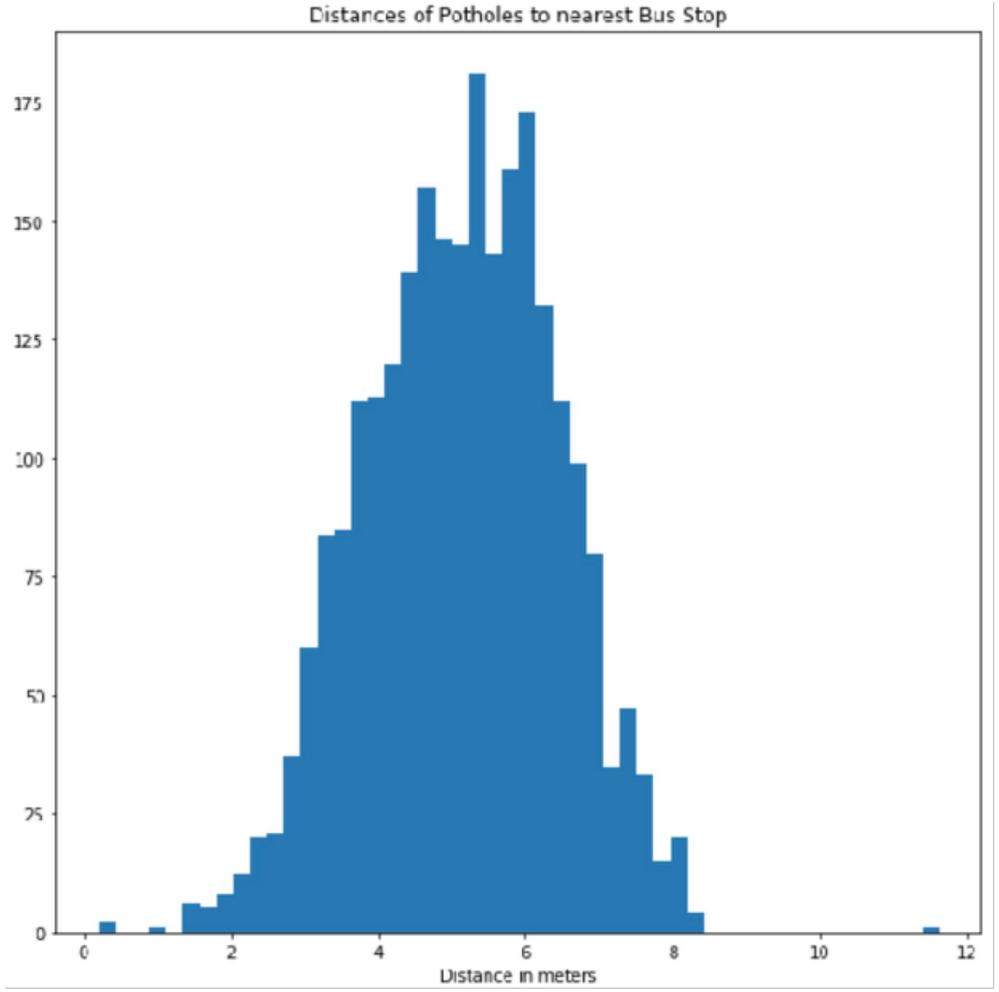
Monte Carlo

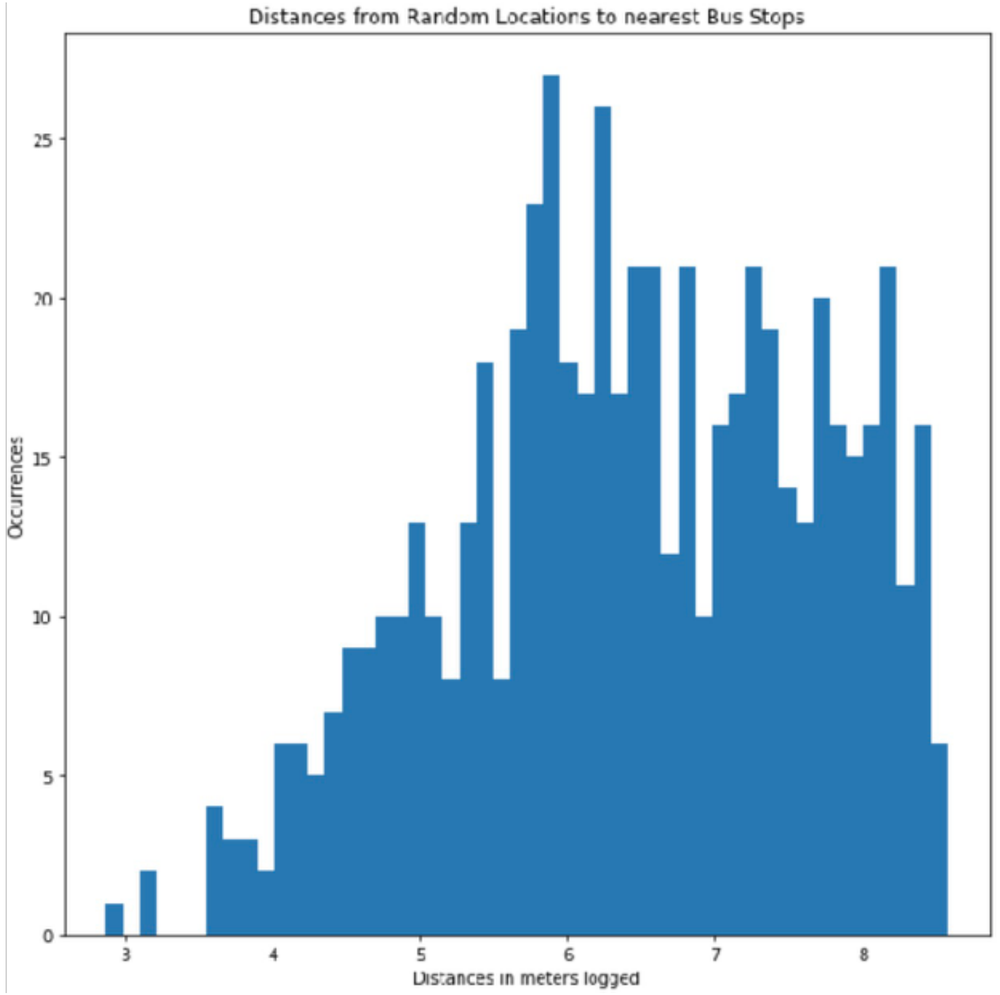


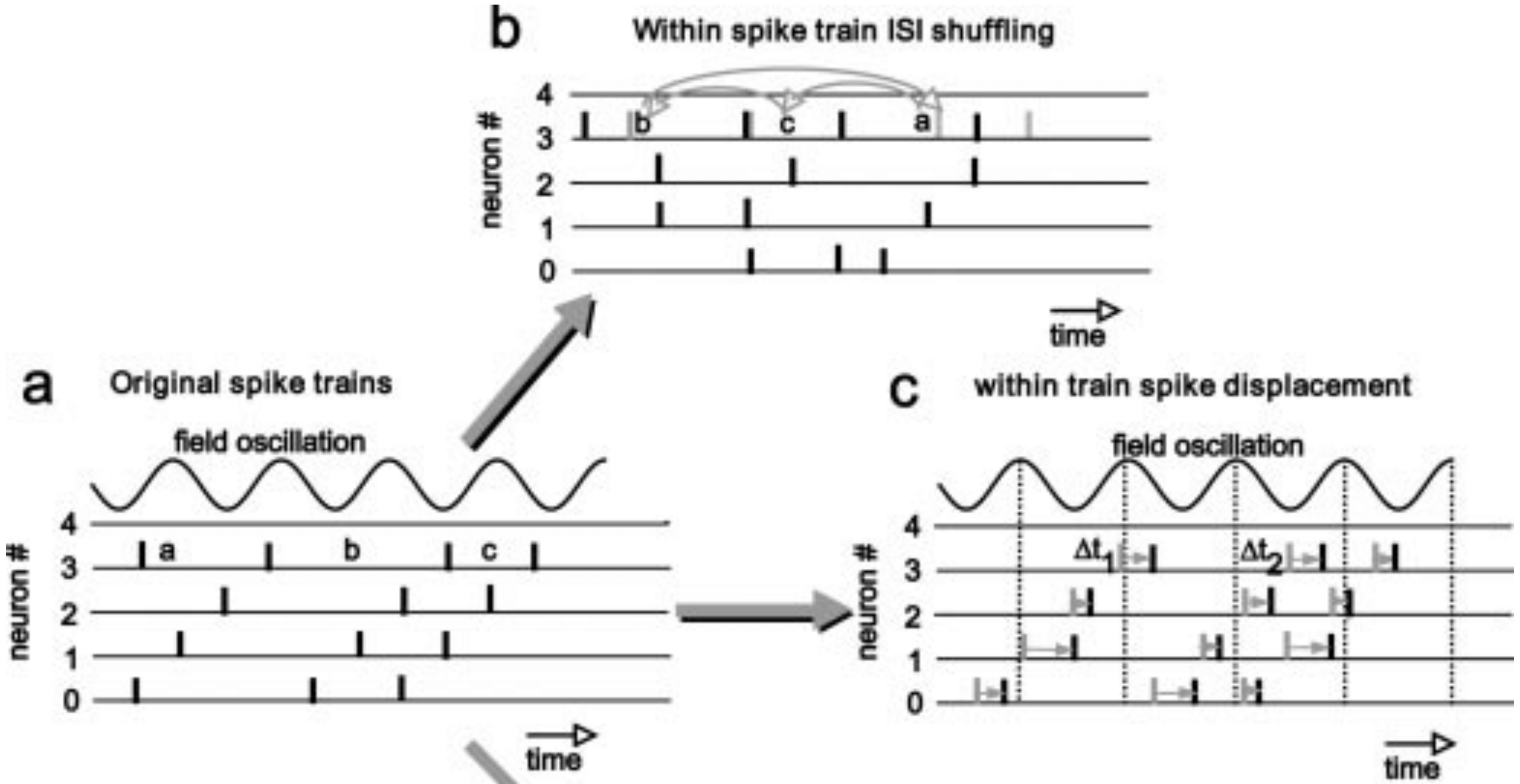
stop



Monte Carlo


Distances of Potholes to nearest Bus Stop
175
150
125
100
75
50
25
0
0
2
4
6
8
10
12
Distance in meters


Distances from Random Locations to nearest Bus Stops
25
20
15
10
5
0
Occurrences
3
4
5
6
7
8
Distances in meters logged


b
Within spike train ISI shuffling
neuron #
4
3
2
1
0
b
c
a
time
a
Original spike trains
field oscillation
neuron #
4
3
2
1
0
a
b
c
time
c
within train spike displacement
field oscillation
neuron #
4
3
2
1
0
Δt1
Δt2
time

Monte Carlo techniques

- Build a null distribution through a simulation of what things would look like under the null through random number generation
- Examples:
  - What if things are spatially random? Throw darts at the map, i.e. draw locations under uniform random sampling. Compare to empirical distribution
  - What if events are at random time? Take inter-event intervals as fixed, but shuffle the intervals. Compare with the timing of empirical data.

2) Rank Statistics

•We rank things in the real world *all the time!*

- International rankings (economics, happiness, government performance)
- Sports (teams, players, leagues)
- Search Engines
- Academic Journals’ prestige
- Reviews online (1-4 stars)

Rank Statistics

quantitative data		ordinal data
1, 4.5, 6.6, 9.2	→	1, 2, 3, 4

Data are transformed from their quantitative value to their rank.

Ordinal data - categorical, where the variables have a natural order

Particularly helpful when data have a ranking but no clear numerical interpretation (i.e. movie reviews)

Ordinality

Which of the following variables contains ordinal data?

https://forms.gle/UUbbQd9nyz8tgzVh6

Wilcoxon rank-sum test (Mann Whitney U test)

- Determine whether two independent samples were selected from the same populations, having the same distribution

- Similar to t-test (but does not require normal distributions) & tests median

•Assumptions:

- Observations in each group are independent of one another

- Responses are ordinal

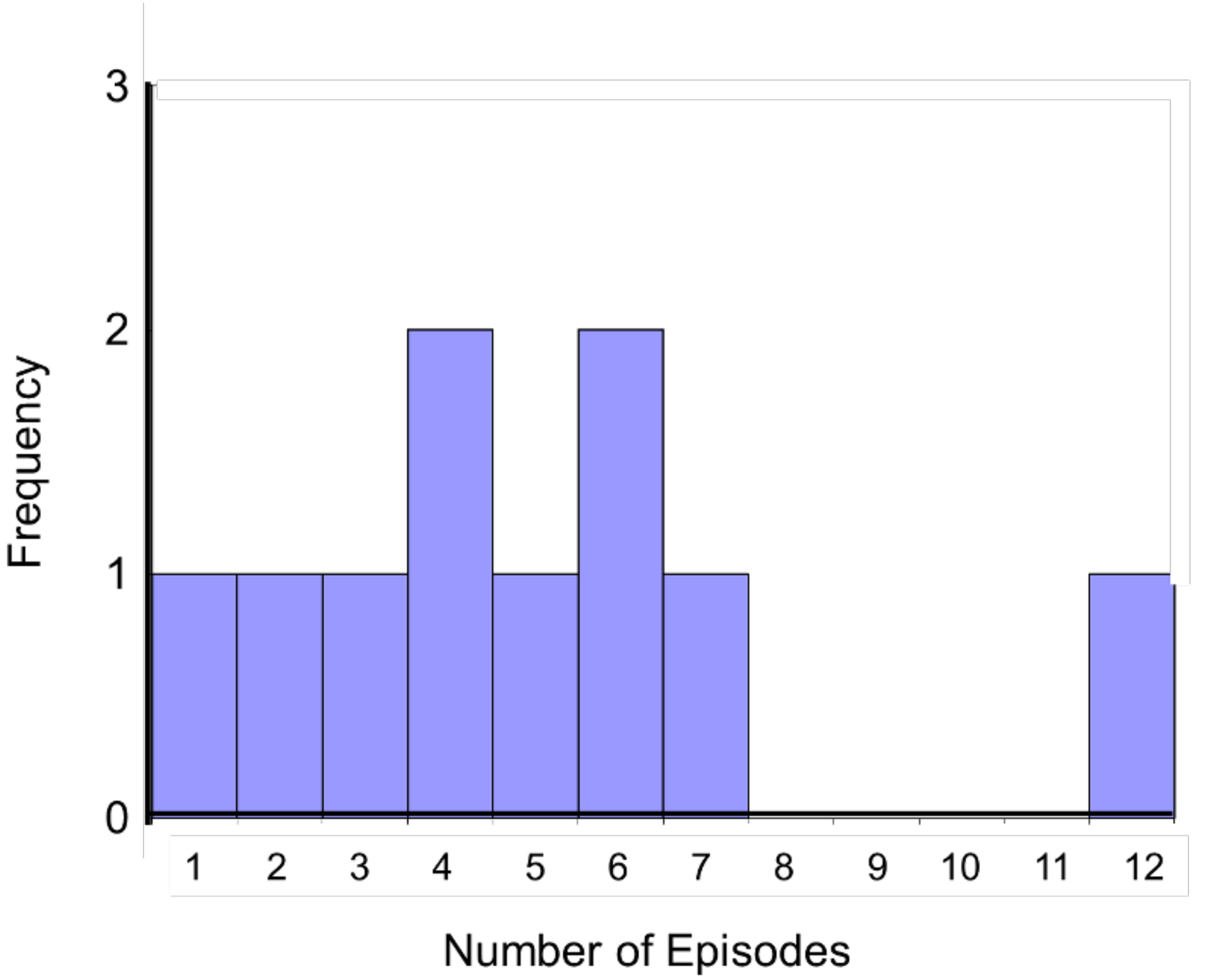
•H_o: distributions of both populations are equal

•H_a: distributions are *not* equal

Mann-Whitney U: question example

- In a clinical trial, is there a difference in the number of episodes of shortness of breath between placebo and treatment?

- Step 1: Participants record number of episodes they have.
- Step 2: Episodes from both groups are combined, sorted, and ranked
- Step 2: Resort the ranks into separate samples (placebo vs. treatment)
- Step 3: Carry out statistical test



http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/BS704_Nonparametric4.html

		Total Sample (Ordered Smallest to Largest)	Ranks
Placebo	New Drug		
7	3		
5	6		
6	4		
4	2		
12	1		

Sum of ranks:
Placebo = 37
New Drug = 18

http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_

Mann-Whitney *U*

Ho: low and high scores are approximately evenly distributed in the two groups

Ha: low and high scores are NOT evenly distributed in the two groups (U <= 2)

$$U_A = n_a n_b + \frac{n_a(n_a+1)}{2} - T_A$$

The max possible value of TA

The observed sum of ranks for sample A

n_a = number of elements in group A

n_b = number of elements in group B

$U_{Placebo} = 3$

$U_{treatment} = 22$

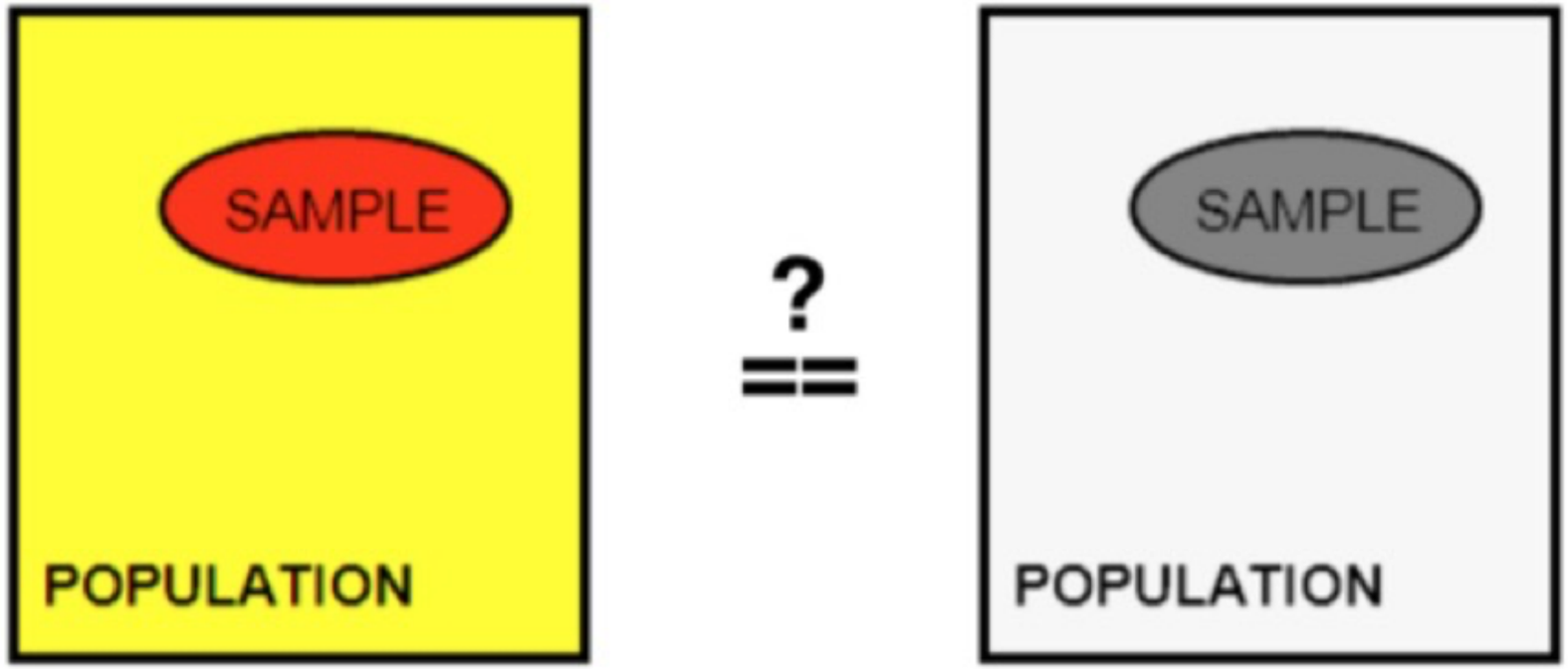
$0 < U < n_1 * n_2$

Complete separation → no separation

We reject the null if U is small.

3) Kolmogorov-Smirnov (KS) test

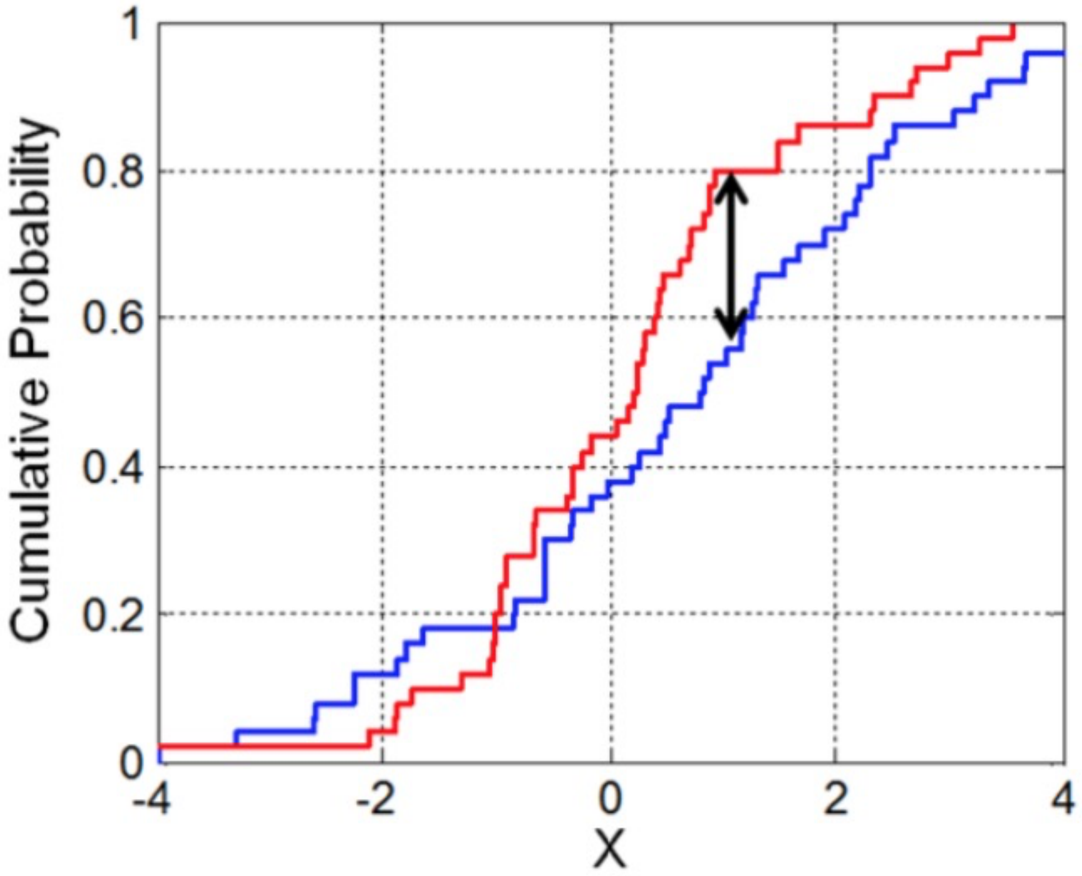
- Given (limited) samples from two populations, how do we quantify whether they come from the same distribution?



Kolmogorov-Smirnov (KS) test

Comparing cumulative distributions empirically

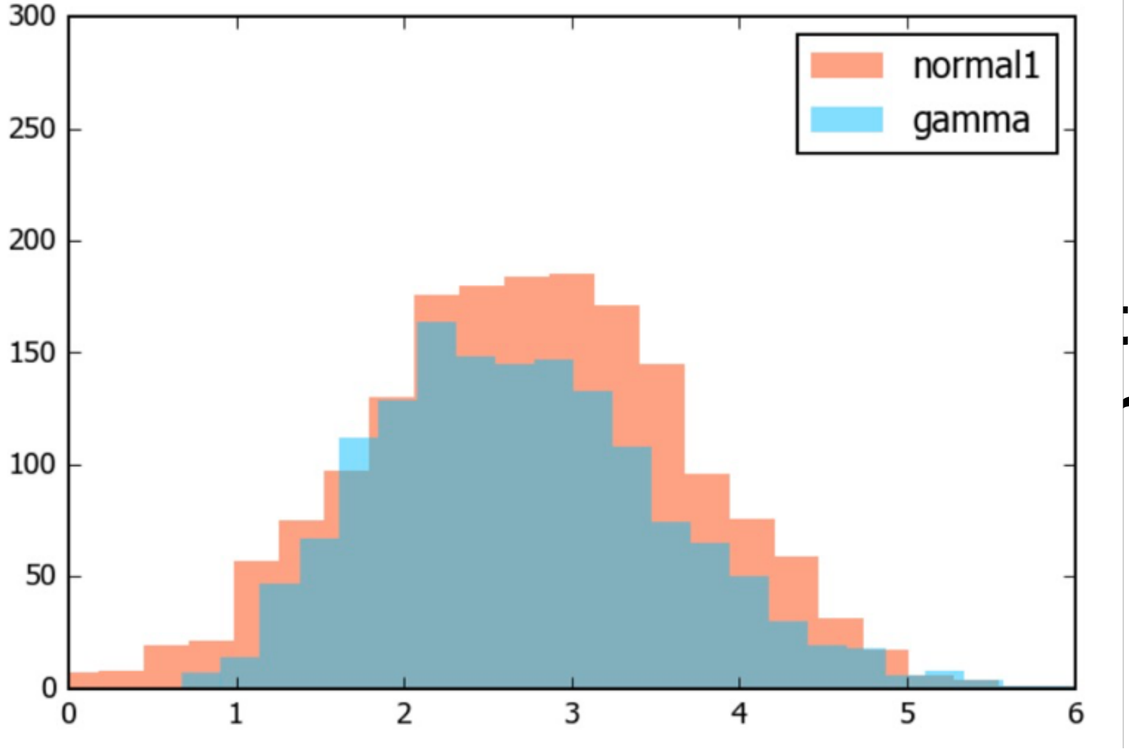
Find the maximum difference between the CDFs.

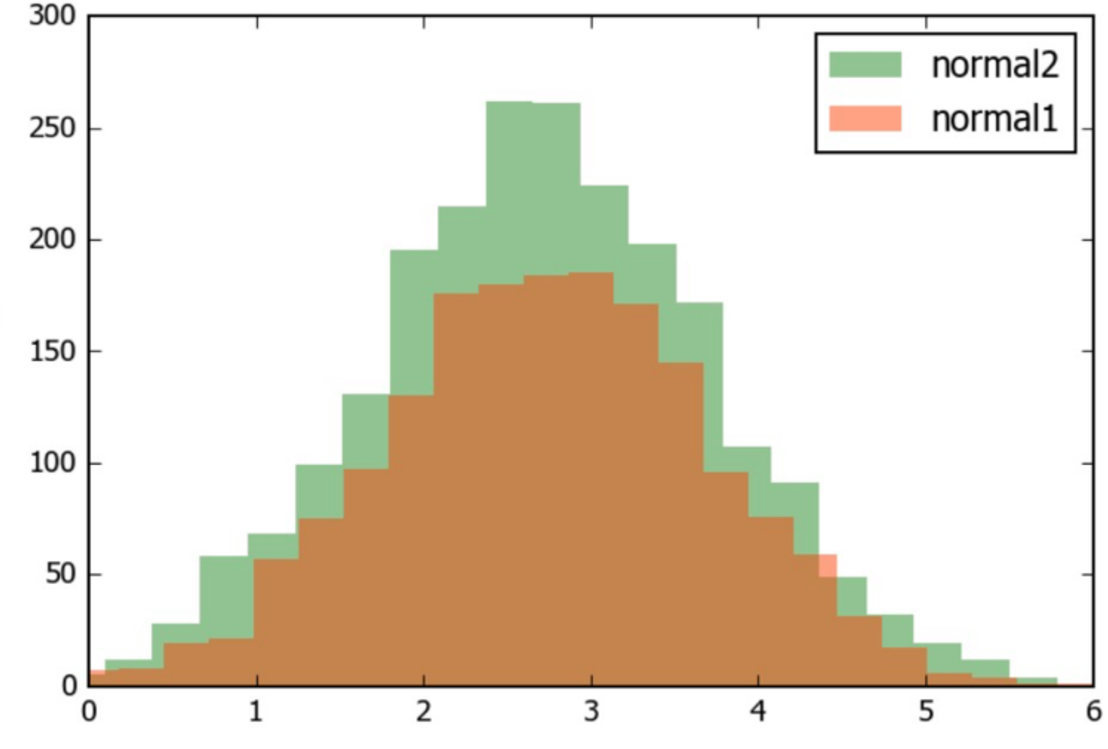


Tests:

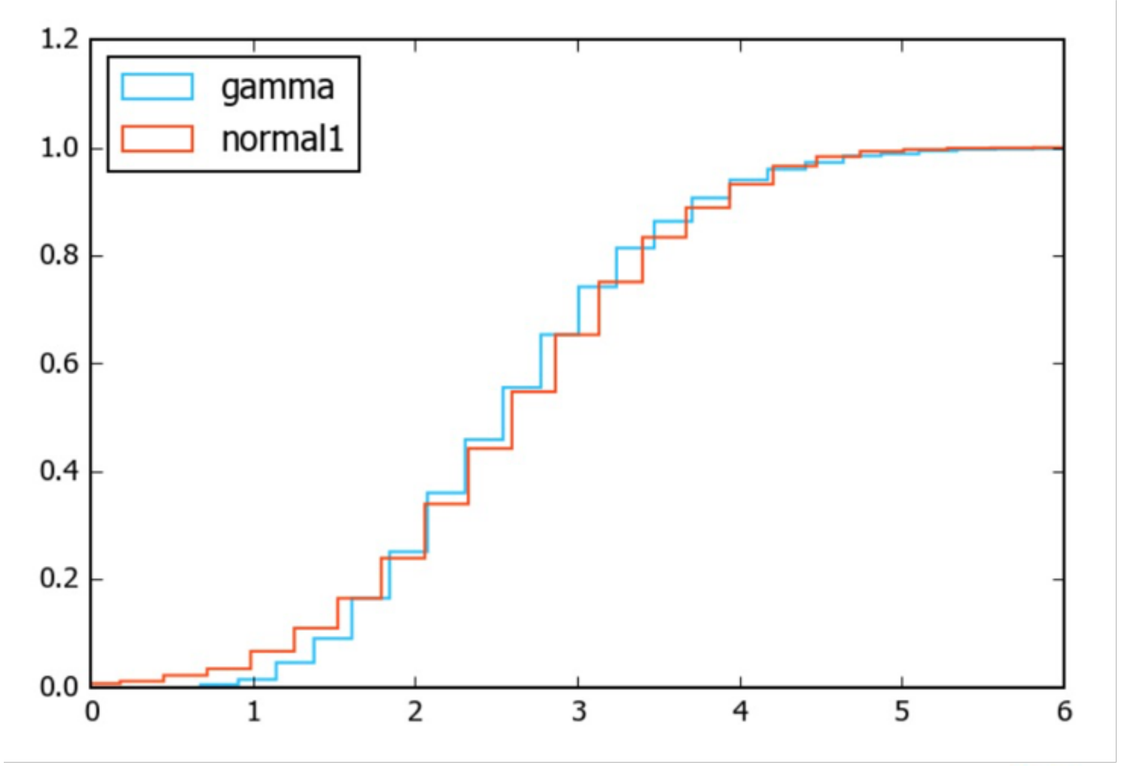
- whether a sample is drawn from a given distribution
- Whether two samples are drawn from the same distribution

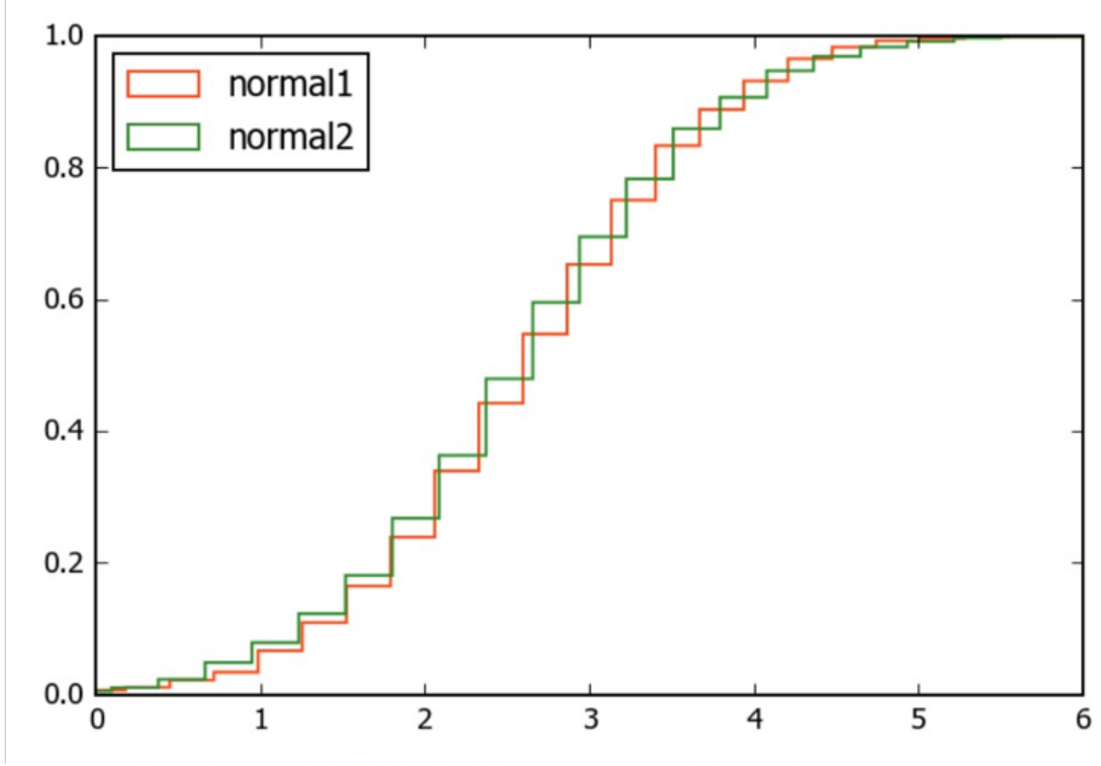
Kolmogorov-Smirnov (KS) test





Kolmogorov-Smirnov (KS) test






```
gamma vs. normal1: p = 0.010680362841 1
normal1 vs. normal2: p = 0.550735998243
```

4) Non-parametric prediction models

- When you have lots of data and no prior knowledge
- When you're not focused/worried about choosing the right features
- Goal: fit training data while being able to generalize to unseen data
- Examples:
  - KNN (K-Nearest Neighbors)
  - Decision Trees and Decision Tree Regressions (CART)
  - Quantile regression

Why do we even teach/use parametric statistics anyway?

- Parametric approaches:
  - Lots of data follow expected patterns
  - Require less data
  - More sensitive
  - Quicker to run/train/predict
  - More resistant to overfitting