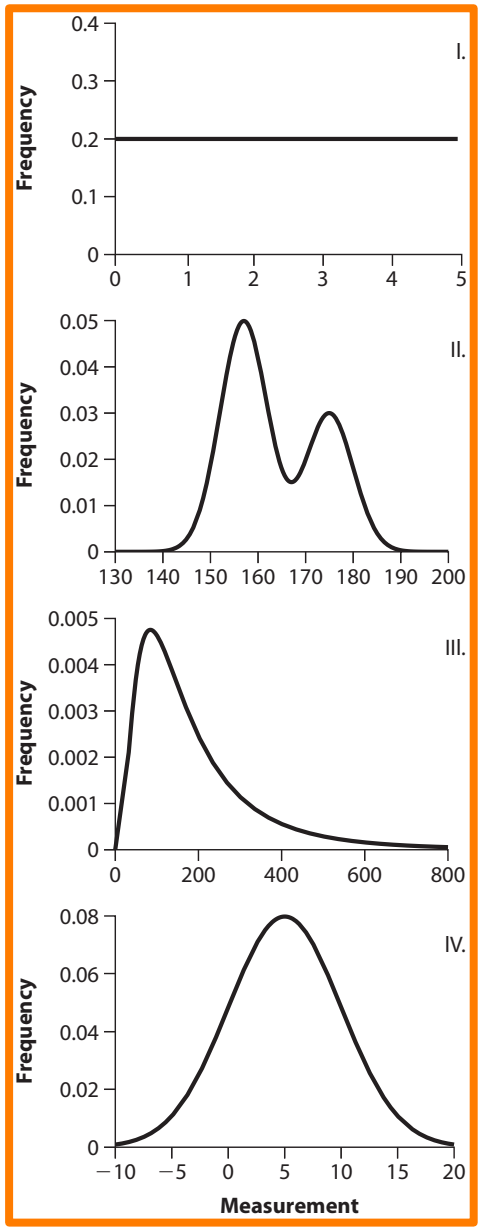
Nonparametric Tests



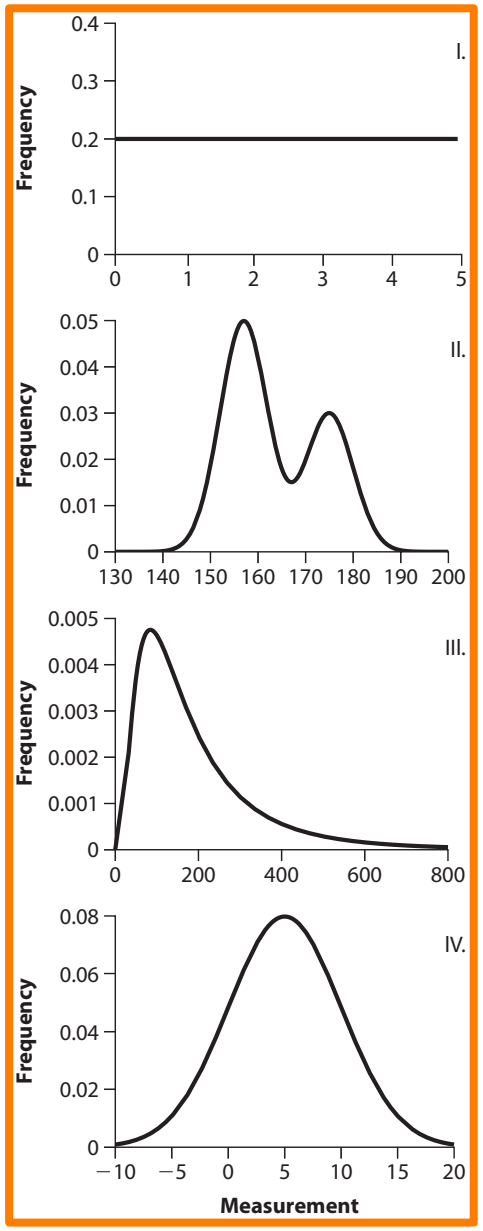
For each graph, what is the best strategy for testing a null hypothesis that the mean or median is 0?

a. Sign test, sign test, transformation/t-test, t-test

b. Sign test, transformation/t-test, sign test, t-test

c. Sign test, transformation/t-test, transformation/t-test, t-test

d. transformation/t-test, transformation/t-test, transformation/t-test, shapiro-wilks test/t-test



For each graph, what is the best strategy for testing a null hypothesis that the mean or median is 0?

a. Sign test, sign test, transformation/t-test, t-test
b. Sign test, transformation/t-test, sign test, t-test
c. Sign test, transformation/t-test, transformation/t-test, t-test
d. transformation/t-test, transformation/t-test, transformation/t-test, shapiro-wilks test/t-test

Answer: "A"

- i – You can't 'transform away' the Uniform distribution
- ii - You can't transform away two 'humps'
- iii – You probably could log transform to get normal distribution, followed by a one sample t-test
- iv – normal distribution allows you to use a one sample t-test automatically.

Benefits of nonparametric methods:

- Assume less about the underlying distribution
- Also called "distribution-free"
  - recall that parametric methods make some sweeping assumptions about distribution
- Particularly useful property with outliers

General concept:

- Most nonparametric tests use **RANKS**
  - data points are not compared, their ranks are!
  - Using **ranks** is what frees us from having to assume normality since all distributions have similar predictions about ranks
- Rank each data point in all samples from lowest to highest
  - Lowest data point gets rank 1, next is rank 2, ..., highest is rank n
- **If H_0 was true than the ranks would be approximately the same between groups**
  - if one group has most small measurements and the other has largest then we may be able to reject equality

Sign Test

- used in place of one sample t-test or paired t-test when normality can't be assumed

- compares data from one sample to a constant (the hypothesized median)

Sign Test

Method:

- for each data point, record whether individual is above (+) or below (-) the hypothesized median

- Use a binomial test to calculate P-value compared to a result of ½
    - if the data points were randomly distributed among the two populations then the median rank in each population would be 1/2

Sign Test

- Throws out a lot of information
- Not very powerful
  - FTR a false H_0

- n > 5*

- Sometimes nothing else works…

* If there is no real difference, a sample size of n=4 will reject the null ~12.5% of the time and n=5 will reject the null ~6.25% of the time (Dixon, 1946). Usually you would deploy the sign test when you have sample sizes that are >> 15. However, that isn't practical in classes like this one where you sometimes have to make calculations by hand. Besides, you can get away with smaller sample sizes if the difference is VERY BIG.

Sign Test

Example: **Is polygamy associated with higher or lower speciation rates?**

- Sexual conflict between males and females may drive genetic divergence between isolated populations of the same species --> formation of a new species!
- Sexual conflict is more pronounced in species where females mate > 1 males

- *25 taxa where females mate > 1 are paired with closely related monogamous species*

Arnqvist et al (2000). Sexual conflict promotes speciation in insects.PNAS 97:10460 -10467

Sign Test

Example: **Is polygamy associated with higher or lower speciation rates?**

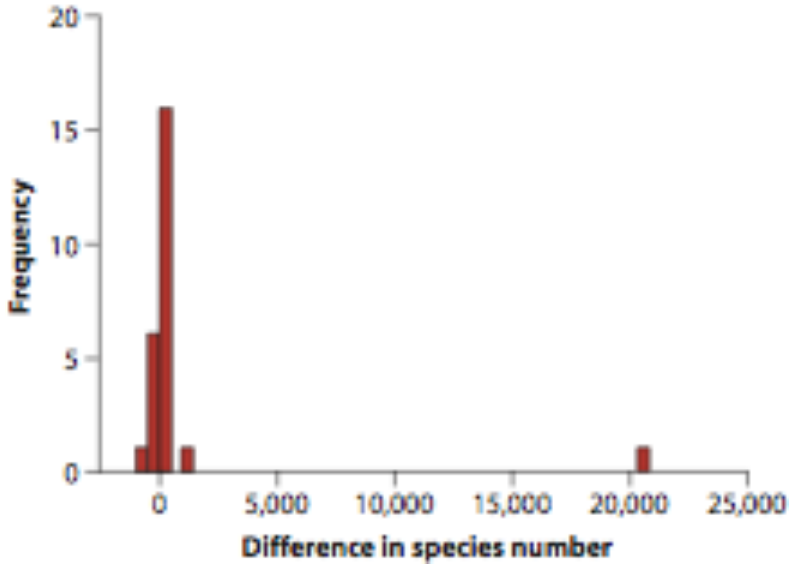
Taxon pair	Multiple Mating	Single Mating 10	Difference	Above (+) Or below (-) Zero
A	53	10	43	+
B	73	120	-47	-
Y	8	8	78	+

- data is paired
  - analysis is for paired differences
  - always keep the same order for calculating differences

Sign Test

Example: **Is polygamy associated with higher or lower speciation rates?**

- Test for normal distribution of differences in species numbers between polygamous and monogamous species



Not normal and too small a sample size… move on

Sign Test

Example: Is polygamy associated with higher or lower speciation rates?

H_0: the median difference in number of species between monogamous and polygamous insect groups is 0

H_A: the median difference in number of species between monogamous and polygamous insect groups is NOT 0

Sign Test

Example: **Is polygamy associated with higher or lower speciation rates?**

"Ranks" of differences:

43 -47 154 64 127 296 16 -100 -980
-290 1090 -8 -78 70 20400 -3 2 8 12
227 1 61 1 79 78

- 18 of the comparisons are above 7 are below
- Use the binomial distribution to calculate P-value

$$P[X \leq 7] = \sum_{i=0}^{7} \binom{25}{i} (0.5)^i (0.5)^{25-i} = 0.02164$$

Sign Test

Example: Is polygamy associated with higher or lower speciation rates?

$$P[X \leq 7] = \sum_{i=0}^{7} \binom{25}{i} (0.5)^i (0.5)^{25-i} = 0.02164$$

This is a two-tailed test so the P-value must be doubled

P-value = 2(0.02164) = 0.043

Since the P-value < α =0.05, we reject the null hypothesis. Groups of insects whose females are polygamous do no produce the same number of species than those whose females are monogamous; this is consistent with the sexual-conflict hypothesis.

Wilcoxon signed-rank test

- similar to sign test

- retains information about magnitudes of distance of points away from hypothesized median

- assumes symmetric distribution with no skew (or outliers)