Modern Statistics Methods

Sampling distributions:

1. Simulation – hypothesis testing

2. Randomization/permutation (technically chapter 13)

Hypothesis testing

Precision of estimates:

3. Bootstrapping

Null distributions:

- Binomial
- Normal
- Poisson
- X^2
- F
- student's t

t-test

ANOVA

Regression

Correlation

X^2 GOF

X^2 Contingency

Mann-Whitney U

Sign test

Kruskall-Wallis test

Spearman

Null sampling distributions:

1. Simulation – hypothesis testing

2. Randomization/Permutation

Precision of estimates:

3. Bootstrapping – sampling distribution of estimate; the values for the parameter estimates that we might obtain and their probabilities. Usually rely on normal approximation for the sampling distribution.

Randomization/Permutation:

- Asks: **are two variables independent?**
- **Assumptions:** random sampling, distribution of variable has approximately same shape

- Versatile
  - Variables can be any combination of numerical or categorical
  - We don't need a null hypothesis because we built it ourselves
  - **MWU test is a type of permutation tests** – but you lose power when you use ranks instead of the actual data
- Basis: **Permutation**
  - Sampling without replacement

Randomization:

Basis: **Permutation**

- Sampling without replacement
- A randomization test generates a ***null distribution*** for the association between two variables

Method:

1. Create data set
   - Response variable of a test statistic measuring association **randomly assigned to Explanatory variable**
     - **You are effectively exchanging labels**
   - **All data points are used exactly once**

2. Calculate measure of association for randomized sample

3. Repeat randomization many times
   - A NULL distribution

Pretty much gives you a p-value and not much else!

Randomization:

The following is a very small data set of birth weights (in kg) of either singleton or individuals who were born with a twin. Create a legitimate randomized data set:

Singleton: 3.5, 2.7, 2.6, 4.4

Twin: 3.4, 4.2, 1.7

Randomization:

The following is a very small data set of birth weights (in kg) of either singleton or individuals who were born with a twin. Create a legitimate randomized data set:

Singleton: 3.5, 2.7, 2.6, 4.4

Twin: 3.4, 4.2, 1.7

Randomization:

Singleton: 1.7, 4.4, 4.2, 2.7

Twin: 2.6, 3.4, 3.5

Of the two scatter plots shown, one represents the real relationship between body size and brain size for 29 species of dinosaur and the other shows the first randomized data set created from that same data.

Which do you think is the real data and why?

A. Top

B. Bottom