Module 4: Drawing Conclusions From Data

4.2 Tests of Significance, Randomization Test, and Bootstrapping

Tests of Significance

- Much of statistical analysis revolves around testing hypotheses
- Recall that a null hypothesis is set up (typically that no effect exists)
- The alternative hypothesis typically is that some effect exists
- Data are then used to calculate a statistic

Tests of Significance

- That statistic is compared against a distribution to see what the probability is that it could have occurred if the null hypothesis is true
- If the probability (p-value) is low (below a set level - typically 0.05 or 0.01) then the null hypothesis is not likely to be true and is rejected

Module 4.2

Tests of Significance

- However, there are some issues with this approach:
  - Small data sets favor accepting the null hypothesis
  - Large data sets favor rejecting the null hypothesis
- So, sample size is important and can even be used unethically to get a desired result

Module 4.2

Tests of Significance

- Confidence intervals use the best estimate of a parameter and of its standard error to give an interval within which the true parameter is likely to lie
- It may be more appropriate and useful to report results using confidence intervals than p-values
- Also use graphics to visually show both data and results!

Tests of Significance, Randomization Test and Bootstrapping

- Another issue involves the assumptions that must made regarding data distributions to allow standard statistical techniques to be used.
- Parametric tests require distributional assumptions, nonparametric methods do not.
- Nonparametric methods are very useful in environmental science!

Tests of Significance, Randomization Test and Bootstrapping

- For example, to carry out a test of whether or not the means of two populations are different:
  - Use the two-sample t test
  - Use Fisher's randomization test
  - Use Bootstrap Methods
  - Use other nonparametric techniques
- The two-sample t test assumes that both population distributions are normal with the same variance
- The nonparametric tests assume nothing

Module 4.2

Randomization Test

- Fisher's randomization test is an example of a nonparametric test - a test that does not require assumptions about the data distribution
- It's easy to perform and intuitive to understand but not often used in practice. The disadvantage of the randomization test is that it is time or computer intensive.
- However, with today's computers, this is not so important.

Module 4.2

Randomization Test

- The randomization test uses resampling as its basis.
- Resampling means that all of the data are thrown into a pool and many new samples are drawn from that pool
- Statistics are then calculated on these new samples (resamples) and these are used as the basis for deciding if the original sample supports the null or alternative hypothesis

Randomization Test

- Example
  - I'll use a very small data set for illustration
  - Three samples from two populations gives 6!/3!*3! or 20 possible arrangements
  - Data:

6	3
5	5
8	4

- Question: Are the population means different?

Randomization Test

- Sample mean of first set = 6.3
- Sample mean of second set = 4.0
- The difference of the means = 2.3
- Now lump all six data points together and randomly pick three for set 1 and three for set 2. Calculate the absolute value of the difference of the means.
- Repeat a large number of times.
- Sort the differences from largest to smallest

Module 4.2

Randomization Test

- Since there are so few possible resamples from only six data points, and they are all equally likely, they can be displayed in a table.
- Normally more data exists and actual random samples are drawn rather than all possibilities being enumerated

Module 4.2

Randomization Test

Set		Resamples with observed mean differences											
1	**6**	6	6	6	6	6	6	3	5	4			
1	**5**	5	5	5	3	5	4	5	5	5			
1	**8**	3	5	4	8	8	8	8	8	8			
2	**3**	8	3	8	5	3	5	6	6	6			
2	**5**	5	8	3	5	5	3	5	3	5			
2	**4**	4	4	5	4	4	5	4	4	3			
d_i	**2.3**	**1.0**	**0.3**	**0.3**	**1.0**	**2.3**	**1.7**	**0.3**	**1.7**	**1.0**			
6	6	6	3	3	5	3	3	5	3				
3	3	5	5	5	5	5	4	4	5				
5	4	4	5	4	4	8	8	8	4				
5	5	5	6	5	6	5	5	5	5				
8	5	8	8	6	8	6	5	6	6				
4	8	3	4	8	3	4	6	3	8				
1.0	**1.7**	**0.3**	**1.7**	**2.3**	**1.0**	**0.3**	**0.3**	**1.0**	**2.3**				

Randomization Test

- The observed difference is equivalent to the largest difference possible in this set of data
- In this case it is in the top 20% of values
- 20% is not low enough to reject the null hypothesis however this is due to the small number of data points

Sorted Diffs	
2.3	1.0
2.3	1.0
2.3	1.0
2.3	1.0
1.7	0.3
1.7	0.3
1.7	0.3
1.7	0.3
1.0	0.3
1.0	0.3

Bootstrapping

- Bootstrapping is another computer intensive technique that uses resampling
- A difference between bootstrapping and randomization is that bootstrapping samples with replacement meaning that one data point could appear two or more times in a resample

Module 4.2

15

Bootstrapping

- For each resample, a sample t statistic is calculated
- These are used to approximate a t distribution specific to this data set
- The critical values are then pulled off of this approximate t distribution (instead of using a table)
- These are used to calculate the confidence interval

Module 4.2

16

Bootstrapping

- To do a Bootstrap t (Manly pages 108-110), you resample from the original data set of n points, m times
- See the Instructions under Help for step-by-step directions
- Once you have the m resamples, calculate the sample mean, sample standard deviation, and sample t for each resample

Bootstrapping

$$t_i = \frac{\overline{X}_i - \overline{X}}{s_i / \sqrt{n}}$$

Calculate t_i where $Xbar_i$ is resample mean, X bar is original sample mean, s_i is resample standard deviation, n is both original sample size and resample size.

Sort the t_i values from smallest to largest and determine the values that have 95% between them and 5% outside of them. See the Practice Problem Set for hints on how to do this and an example when m=50.

Bootstrapping

- The confidence interval is equation 4.3 in Manly, page 110.
- (Xbar – t_{high}*s/sqrt(n), Xbar – t_{low}*s/sqrt(n))
- Note these are both minus, the t_{low} value will be negative so the double negative becomes a plus.
- This is your confidence interval without using the assumptions of the CLT

Bootstrapping

- A Bootstrap CI does not have to be symmetric, in fact, when they are useful they are generally not symmetric
- Why?
- The sampling distribution of the sample mean approaches normal as n gets large but is asymmetric for small sample sizes from skewed distributions

Bootstrapping

- So, if the data distribution is skewed and n is small, the sampling distribution of the sample mean will be skewed and a confidence interval on the mean should be asymmetric.
- The Bootstrap captures this through simulation.