

Permutation Tests & Bootstrapping

a Practical Stats Online Course

Dennis R. Helsel, Instructor

May 10-24, 2016

May 10, 2016

1. Basics of Permutation Tests

- What is a permutation test?
- Why are permutation tests important to environmental science?
- How do permutation tests produce a p-value?
- What are the advantages over parametric test methods?
- Comparing permutation tests to the t-test and Wilcoxon rank-sum test
- Useful references for permutation tests

2. Use of R and RStudio

- Software for permutation tests
- Example: a two-group permutation test using R software
- The homework assignment

May 17, 2016

1. Paired and Unpaired Permutation Tests

- Computing the two-group permutation tests - review
- Matched-pair tests

2. Permutation Tests for Three or More Groups

- Comparing permutation tests to Analysis of Variance and Kruskal-Wallis
- Computing the one-factor permutation test
- Two-factor permutation tests vs. ANOVA
- The homework assignment

May 24, 2016

3. Bootstrap Confidence Intervals

- What is bootstrapping?
- Bootstrap intervals for the mean
- Advantages over t-intervals for skewed data
- What parameters can be bootstrapped?
- Example: bootstrap estimate of the UCL95 for background data
- The homework assignment

Permutation Test "flowchart"
 PracticalStats.com
 January 2016

Objective	Parametric Method (<i>outdated</i>)	Permutation /Bootstrap Commands in R
Conf interval on Mean	t-interval	boot.ci
UCL	t-interval	boot.ci
LCL	t-interval	boot.ci
Tolerance interval on percentile	parametric tolerance interval	bootTOLupper *
Test difference in 2 groups' means	two-sample (or "independent samples") t-test	permTS
test mean=Value	one-sample t-test	perm1sample *
Test mean for paired samples	paired sample t-test	permMatched *
Test mean of 3+ groups	ANOVA	permKS
Two-factor ANOVA	Factorial ANOVA	independence_test

* not natively in R. These are PracticalStats.com scripts, © Practical Stats 2016. For use only by students who have taken this class.

Instructions for installing R for the Practical Stats Online Course
Permutation Tests and Bootstrapping
May 2016

Instructions for installing R. For information on system requirements, etc. go to the R FAQs page, at

<http://cran.r-project.org/bin/windows/base/rw-FAQ.html>

1. **Install R.** Go to <http://cran.r-project.org/> . In the top box titled “Download and Install R”,
 - a) select “Download R for Windows”. Then select “base” to install the Rgui.exe program. Use the default settings. OR
 - b) select "Download R for (Mac) OSX". Click on the first entry, "R-3.3.0.pkg" to install.

2. **Download and install RStudio**

Go to www.rstudio.com and click the Download RStudio button. Follow the instructions.

3. **Install 4 additional support packages.** R is modular – you install what you need.

- a. Start up RStudio.
- b. Go to the Packages tab (in the lower right window) and click "Install". In the new window, make sure the "Install Dependencies" box at the bottom is checked (it probably is). Then type in one of the following package names:

boot car coin perm

and click the Install button. This will place that package onto your hard drive. Repeat until all four are installed (added to the list of packages in the Packages tab).


Email me with any questions: dhelsel@practicalstats.com

Permutation Tests and Bootstrapping

"Homework" Exercise

May 10, 2016

(Complete before the next session)

1. Follow the instructions on the previous page and install R and support packages, and RStudio, onto your computer. First talk it over with your IT staff.
2. Start RStudio. Load (activate) the four packages by clicking the boxes next to their names in the Packages list. If they aren't listed in the Packages list, go back to step 3 in the instructions on the previous page and install them.
3. Set your working directory. Go to Session > Set Working Directory > Choose Directory and find the PermBoot Data directory you previously downloaded for this class. Select it.
4. Load the MOLY2 dataset. Go to the Environment tab in the upper right window and click on its OpenFile icon . If you then click on the dataset name that appears there, the dataset will show in a tab in the top left window.
5. In the top left window, click its OpenFile icon. Load the "PermOnline Commands May 10.R" file from the PermBoot Data directory. These are commands that if done one by one will complete the exercise. If you're an experienced R user and don't need these, ignore them. Control clicking anywhere in the command line will run that line in the console (lower left) window.
6. Complete the exercise described on the next page. Draw a boxplot and test differences in the mean molybdenum concentrations for the two groups (DOWNGRAD vs upgrad) of MOLY2 data. Run both the t-test and the permutation test (permTS command).

Testing differences between two independent (unpaired) groups

A shallow aquifer is contaminated by leachate from mine tailings. One of the constituents is molybdenum. A large irrigation ditch crosses downgradient of the contamination site, allowing low-molybdenum concentration surface waters to infiltrate into the aquifer.

Data from the 3 upgradient wells and the 13 downgradient wells are found in the dataset MOLY2.

```
print (MOLY2)
      MOLY    LOCAT
1  0.850 DOWNGRAD
2  0.390 DOWNGRAD
3  0.320 DOWNGRAD
4  0.300 DOWNGRAD
5  0.300 DOWNGRAD
6  0.205 DOWNGRAD
7  0.200 DOWNGRAD
8  0.200 DOWNGRAD
9  0.140 DOWNGRAD
10 0.140 DOWNGRAD
11 0.090 DOWNGRAD
12 0.046 DOWNGRAD
13 0.035 DOWNGRAD
14 6.900  upgrad
15 3.200  upgrad
16 1.700  upgrad
```

Test whether the aquifer downgradient of the ditch is significantly different in molybdenum than is the upgradient.

Try appropriate parametric and permutation tests.

Homework Assignments: Session 2

Testing differences between two independent groups

Dissolved oxygen was measured in a river/estuary in Florida over several years. In 2008 a change was made upstream that was hoped to increase DO conditions somewhat. However, the uncertainty in the effect was enough to test also for a decrease in DO – the scientist did not want ignore a decrease if it was observed (a one-sided test for increased DO would ignore any decrease). Therefore, run a two-sided test for whether DO has changed.

1. Plot the data first.
2. Run a t-test to determine whether the mean has changed. Don't forget to also test for normality of the residuals.
3. Run a wilcoxon rank-sum test to determine whether the median has changed.
4. Run the 2-sample permutation test to see if results differ from the t-test (it is the main subject of this class, after all).
5. If the interest is in whether DO is typically higher now than it was prior to the change upstream, what is your answer?
6. If the interest is in whether the cumulative amount of oxygen is higher now than it was prior to the change upstream, what is your answer?
7. What is the primary difference in DO between the two time periods?

Comparing 3+ Groups

Chloride concentrations were measured by Feth et al. (1964) -- *Sources of mineral constituents in water from granitic rocks, Sierra Nevada, California and Nevada; USGS Water Supply Paper 1535-I* – in shallow ephemeral springs and waters from two granitic rock types in the Sierra Nevada mountains. This was a classic paper in geochemistry, not to mention that field work must have been done in awesome scenery. Test whether the mean chloride concentration differs among the groups of springs using anova and the permutation test. Test whether median concentrations differ using the nonparametric analog. Explain your findings.

Testing Differences in Two Paired Groups

Total Phosphorus was measured at three sites below the impoundment on the West Branch of the Penobscot River. Compare two of these downstream sites, sites DP2 and DP3, to see if Total P is the same at both. What characteristic is used to pair data between the groups? Compute the paired t-test and the paired permutation test and interpret the results. The data are in file DP2v3_TP.Rdata