Lab 10: nonparametric inference

STAT218

The goal of this lab is to learn how and when to implement nonparametric alternatives to one- and two-sample tests. There are three alternatives we will consider:

* signed rank test (one-sample or paired differences)
* rank sum test (two-sample inference for independent data)
* permutation test (two-sample inference for independent data)

We will illustrate the implementation of these tests using the following datasets:

* ddt contains 15 measurements of DDT in kale
* sleep contains 10 paired measurements of extra hours of sleep on two different drugs
* cancer contains observations of cancer rates in CT and whether sunspot activity was higher or lower than average in each of 35 years, along with the deviation of the cancer rate from expectation each year

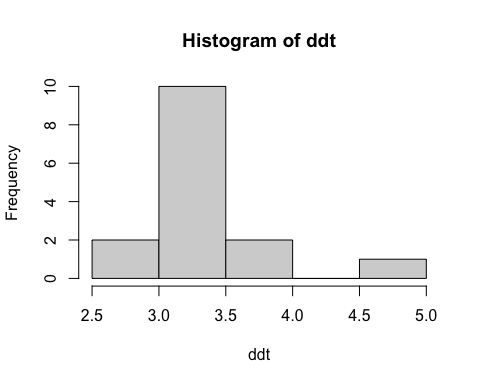
# load example datasets  
ddt <- MASS::DDT  
sleep <- read.csv('data/sleep.csv')  
load('data/cancer.RData')

To determine whether the rank procedures are appropriate, histograms should be inspected for:

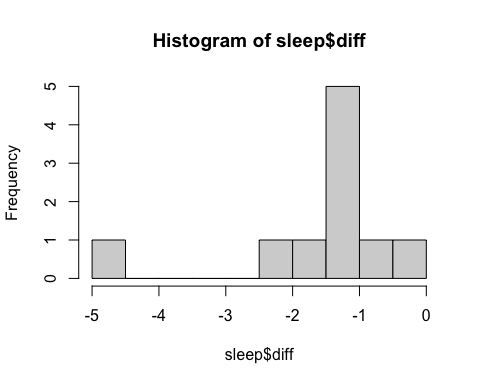
* symmetry in the one-sample or paired inference case
* group similarity in the two-sample case

Examples of these histograms are shown below for each dataset.

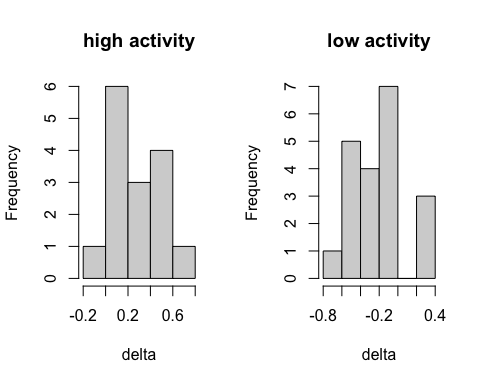
# for one-sample inference, check histogram for symmetry  
hist(ddt, breaks = 5)



# for paired inference, check histogram \*of differences\* for symmetry  
hist(sleep$diff, breaks = 10)



# for two-sample inference, check groupwise histograms for similar shape  
par(mfrow = c(1, 2))  
hist(cancer$delta[cancer$sunspot.activity == 'High'], main = 'high activity', xlab = 'delta')  
hist(cancer$delta[cancer$sunspot.activity == 'Low'], main = 'low activity', xlab = 'delta')



Notice that due to the small sample sizes, symmetry/similarity is hard to assess. As a result, you should only look for *obviously* asymmetric/different shapes. Each of these is acceptable.

The signed rank test and rank sum tests are implemented using wilcox.test().

# signed rank test  
wilcox.test(ddt, mu = 3, alternative = 'greater')

Wilcoxon signed rank test with continuity correction  
  
data: ddt  
V = 111.5, p-value = 0.001876  
alternative hypothesis: true location is greater than 3

# signed rank test, paired differences  
wilcox.test(sleep$diff, mu = 0, alternative = 'less')

Wilcoxon signed rank test with continuity correction  
  
data: sleep$diff  
V = 0, p-value = 0.004545  
alternative hypothesis: true location is less than 0

# rank sum test  
wilcox.test(delta ~ sunspot.activity, data = cancer, alternative = 'greater')

Wilcoxon rank sum exact test  
  
data: delta by sunspot.activity  
W = 270, p-value = 8.182e-06  
alternative hypothesis: true location shift is greater than 0

Permutation tests, by contrast, are implemented using permTS(). Note that these are only applicable for two-sample inference with independent (*i.e.*, not paired) data.

# permutation test  
library(perm)  
permTS(delta ~ sunspot.activity, data = cancer, alternative = 'greater')

Permutation Test using Asymptotic Approximation  
  
data: delta by sunspot.activity  
Z = 3.8143, p-value = 6.829e-05  
alternative hypothesis: true mean sunspot.activity=High - mean sunspot.activity=Low is greater than 0  
sample estimates:  
mean sunspot.activity=High - mean sunspot.activity=Low   
 0.4433333

Permutation tests provide an alternative nonparametric inference procedure with minimal assumptions.

## Sleep data revisited

To start, let’s extend the sleep data example above by trying something a little different than usual. Notice that the estimated difference in means is -1.58, and every subject experienced more extra sleep on drug 2.

Can you work out how to test, using a rank procedure, whether drug 2 is associated with at least one more hour of extra sleep than drug 1?

Write the hypotheses, perform the test, and interpret the result.

# use rank procedure to test whether difference exceeds 1 hour

This is not a scenario specific to nonparametric methods, but just an exercise in testing a less typical set of hypotheses with paired data.

Once you’ve determined and performed your test, write a short report interpreting the result.

## Cholesterol and cereal

Your goal is to test whether cereal is associated with a difference in cholesterol using a nonparametric method.

Use an appropriate graphic to check whether the assumptions for a rank procedure are appropriate; if so, use the appropriate rank procedure; if not, use a permutation test.

# load cholesterol data  
cholesterol <- read.csv('data/cholesterol.csv')  
  
# check assumptions  
  
# determine and perform appropriate test

Once you’ve determined and performed your test, write a short report interpreting the result.

## Zinc and dietary supplements

Your goal is to test whether taking a dietary supplement lowers zinc concentrations among rats using the following data.

# load cholesterol data  
zinc <- Sleuth3::ex0125  
head(zinc, 3)

Group Zinc  
1 A 1.31  
2 A 1.45  
3 A 1.12

The data come from an experiment in which 39 rats were randomly assigned a dietary supplement (group A) or no dietary supplement (group B). After a period of time, the zinc concentration in each rat’s blood was measured.

Use an appropriate graphic to check whether the assumptions for a rank procedure are appropriate; if so, use the appropriate rank procedure; if not, use a permutation test.

# check assumptions  
  
# determine and perform appropriate test

Once you’ve determined and performed your test, write a short report interpreting the result.