Outline

Permutation Bootstrap
  Permutation Tests

Multi-sample tests
  Multi-sample test for location: ANOVA statistic

MANOVA statistics
  MANOVA statistics
  Multivariate permutation tests

References

Exercises
  Mantel test
Permutation Tests for Two Samples

- Permutation tests are based on resampling without replacement.
- Often applied as a nonparametric test of the general hypothesis

\[ H_0 : F = G \quad \text{vs} \quad H_1 : F \neq G, \]  

(1)

where \( F \) and \( G \) are two unspecified distributions.

- Applications include:
  - Tests for location or common scale
  - Nonparametric tests of independence or association
  - Multisample “F” tests for common means or equal distributions (ANOVA, MANOVA)
Under $H_0 : F = G$, two samples from $F$ and $G$, and the pooled sample, are all random samples from the same distribution $F$.

$T$ is a statistic that compares the distributions.

Replicates of $T$ are generated by resampling without replacement from the pooled sample. (a permutation)

The ECDF of the replicates of $T$ is our estimate of the null distribution $T$. 
Permutation Lemma: Each of the replicates is equally likely under the null hypothesis.

If we generate all unique permutations of group labels, the corresponding statistics comprise the permutation distribution.

In this case the permutation test is exact.

Usually not practical to generate all permutations; a large number of randomly generated permutations will suffice.
Multi-sample test for location

- $k$ independent samples are observed.
- Test $H_0 : \mu_1 = \cdots = \mu_k$ without the usual normal theory assumptions about the model.
- Can use the ANOVA $F$ statistic in a permutation test.
- but it is then a nonparametric test (not using $F$ distribution).
- Treatment levels are the group labels.
- Generate permutations of the group labels.
Example chickwts data

chickwts data: the response is weight of baby chicks fed on different types of feed

```r
> with(chickwts,
      by(weight, feed, mean))
```

feed: casein
[1] 323.6

feed: horsebean
[1] 160.2

feed: linseed
[1] 218.8

feed: meatmeal
[1] 276.9

feed: soybean
[1] 246.4

We want to test $H_0$: mean weight is equal for all feed types.
Example chickwts data boxplots
Example 2. chickwts data violin plot
Code for the plots using ggplot2:

```r
> library(ggplot2)
> ggplot(chickwts, aes(x=feed, y=weight)) +
   geom_boxplot(outlier.size = 1.5) +
   stat_summary(fun.y="mean", geom="point", shape=21,
               size=3, fill="blue")
> ggplot(chickwts, aes(x=feed, y=weight)) +
   geom_violin(trim=FALSE) +
   stat_summary(fun.y=mean, geom="point", fill="blue")
```
Randomization of group labels

```r
> table(chickwts$feed)

    casein horsebean linseed meatmeal soybean sunflower
   12     10    12    11    14    12

> as.integer(chickwts$feed)  #original labels

[1]  2  2  2  2  2  2  2  2  2  2  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  5  5  5  5  5  5  5  5  5

> as.integer(sample(chickwts$feed))  #randomized labels

[1]  3  3  4  2  5  2  6  4  2  2  5  3  3  4  3  5  5  2  5  2  1  5  3  5  6  5  1
[39]  6  3  1  2  6  6  4  5  6  6  1  6  1  4  4  2  3  1  4  1  2  5  1  4  3  6  3
```
Example 2. Replicates of ANOVA $F$ statistic

The `oneway.test` function computes the one way ANOVA $F$ statistic (and $F$ test, although we don’t use it here).

```r
> R <- 999
> t0 <- oneway.test(weight ~ feed, 
                   data=chickwts)$statistic
> t0

  F
19.66

> tr <- replicate(R, expr = {
                   rand.feed <- sample(chickwts$feed)
                   oneway.test(weight ~ rand.feed, 
                               data=chickwts)$statistic })
```
Permutation replicates of $F$ statistics

The observed test statistic is 19.66, clearly significant.

> MASS::truehist(tr)
MANOVA-like permutation tests

- MANOVA test statistics can be extracted using the `summary` method of the `manova` object.
- When the usual MANOVA conditions do not hold, we can apply a permutation test.
- We use the `iris` data to illustrate.
- First, let's see the parametric version.

```r
> y <- as.matrix(iris[, 1:4])  # 4 dimensional response  
> species <- iris[, 5]  
> M <- manova(lm(y ~ species))  
> summary(M, test="Pillai")

Df Pillai approx F num Df den Df Pr(>F)
species 2 1.19 53.5 8 290 <2e-16 ***
Residuals 147
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
Easy to extract the MANOVA test statistics

```r
> M <- manova(lm(y ~ species))
> S <- summary(M, test="Pillai")$stat
> W <- summary(M, test="Wilks")$stat
> S

          Df Pillai   approx F num Df den Df Pr(>F)
species    2 1.192     53.47     8 290 9.742e-53
Residuals 147  NA       NA     NA  NA  NA       NA

> W

          Df  Wilks   approx F num Df den Df Pr(>F)
species    2 0.02344 199.10000     8 288 9.365e-112
Residuals 147  NA       NA     NA  NA  NA       NA
```
Extracting MANOVA approx. $F$ statistics

Extract the approximate $F$ statistic for the Pillai test:

```r
> S <- summary(M, test="Pillai")$stat
> S["species", "approx F"]

[1] 53.47
```

Extract the approximate $F$ statistic for the Wilks test:

```r
> W <- summary(M, test="Wilks")$stat
> W["species", "approx F"]

[1] 199.1
```
Multivariate permutation tests

Apply the same algorithm as the multisample location test or multisample test for equal distributions:

▶ Randomize the group labels.
▶ Compute the MANOVA or other multivariate test statistic.
▶ Compare the observed statistic to the replicates to obtain the test decision.

For another example, see the multivariate test for equal distributions in energy package (disco test = DIStance COmponents test).
A nonparametric test based on MANOVA statistics

```r
> n <- NROW(y)
> R <- 999
> t0 <- S["species", "approx F"]
> out <- replicate(R, expr=
    {rand.species <- species[sample(1:n)]
     stats <- summary(manova(lm(y ~ rand.species)),
                      test="Pillai")$stat
     stats["rand.species", "approx F"]
    })
> pval <- (1 + sum(out > t0)) / (1 + R)
> pval

[1] 0.001
```
References


Exercise 1: Mantel test

- Let $X$ and $Y$ be random vectors, and $A$ and $B$ be symmetric distance or dissimilarity matrices for samples from $X$ and $Y$, resp.
- Test for linear association in distances.
- The Mantel $r$ statistic is the Pearson correlation between the $n(n-1)/2$ entries on the upper triangles of $A$ and $B$.
- Usual $t$-test not applicable (distances are not iid).
- Implement a permutation test for test decision.
- Usual method is to randomize the indices of the first sample.
Permutation test for Mantel $r$

- Implement a permutation test for test decision.
- Usual method is to randomize the indices of the first sample.
- Note: not necessary to re-compute distances.
- If $i$ is a permutation of sample indices $1:n$, then $A[i,i]$ are the corresponding distances.
Mantel tests - to check solution

- Try implementing the Mantel $r$ test and/or the partial Mantel $r$ test.
- Check your result by comparison with an available package:
  - `mantel` function in `ecodist` package (for both tests).
  - `mantel`, `mantel.partial` functions in `vegan` package.
  - Note that `ecodist` version is considerably faster than `vegan` version.
  - Tip: To avoid name conflict, use `vegan::mantel` or `ecodist::mantel`. 