Resampling Methods in R

Inga Schwabe
Second TRUG meeting
July 11, 2013
Resampling

- Population $\rightarrow$ Sample $\rightarrow$ Inference
- Classical statistical methods largely based on idealized assumptions (e.g. normal distribution)

Resampling

- Resample from original sample
- No information on population characteristics or distribution

*No assumption on distribution of the data*

- Estimation (e.g. bias), hypothesis testing, model assessment
Resampling (Rodgers, 1999)

<table>
<thead>
<tr>
<th>Theoretical sampling distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Approximation by theoretical distribution (e.g. t, F)</td>
</tr>
<tr>
<td>• Assumptions</td>
</tr>
</tbody>
</table>

v.s.

<table>
<thead>
<tr>
<th>Empirical distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 1 single sample</td>
</tr>
<tr>
<td>• Re-order elements to generate distribution of a statistic</td>
</tr>
</tbody>
</table>
Taxonomy of resampling methods (Rodgers, 1999)

<table>
<thead>
<tr>
<th></th>
<th>With replacement</th>
<th>Without replacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td>Bootstrap</td>
<td>Permutation test</td>
</tr>
<tr>
<td>Subsample</td>
<td></td>
<td>Jacknife, cross-validation</td>
</tr>
</tbody>
</table>
Example permutation test

- Clinical design, treatment and control group

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>2</th>
<th>5</th>
<th>5</th>
<th>6</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

- Was the treatment successful?
- Idea of permutation test: Create a null distribution by random permutation of the data
The steps of a permutation test

- I. Calculate $T$ and p value observed sample
- II. Permutation part
  - (i) Sample without replacement $n$ (16) observations, assign the first 8 observations to the treatment group and the remaining observations to the control group
  - (ii) Compute the $T$ value
  - (iii) Store the $T$ value in a vector/matrix and repeat steps 1-3 $P$ times
- III. Calculate empirical p value
  
  How often $T(\text{permutation}) > \text{observed. } T$ value?
Permutation.T.test <- function (NumSim, group1, group2, side = "two-sided", variance = "equal variances assumed") {

#I. Create variables
n1 <- length(group1)
n2 <- length(group2)
perm_t_values <- matrix(0, NumSim, 1)
degrees_of_freedoms <- (n1+n2)-2
I. Calculate T and p value observed sample

```r
if(variance=="equal variances assumed")
  t_value_obs <- as.numeric(t.test(group1, group2,
    var.equal = TRUE)$statistic)

if(variance=="equal variances not assumed")
  t_value_obs <- as.numeric(t.test(group1, group2,
    var.equal = FALSE)$statistic)
```
I. Calculate $T$ and $p$ value observed sample

```r
if(side=="one-sided") p_value_obs <- 1-qt(t_value_obs,
df = degrees_of_freedom)

if(side=="two-sided") p_value_obs <- 2*(1-qt(t_value_obs,
df = degrees_of_freedom))
```
II. Calculate T and p values permutations

for(i in 1:NumSim){
    permsample <- sample(c(group1, group2), replace = FALSE)
    perm1 <- permsample[seq(group1)]
    perm2 <- permsample[-seq(group1)]

    if(variance == "equal variances assumed")
        t_new <- t.test(perm1, perm2, var.equal = TRUE)$statistic
    if(variance == "equal variances not assumed")
        t_new <- t.test(perm1, perm2, var.equal = FALSE)$statistic

    perm_t_values[i] <- t_new
}

III. Calculate empirical p value

```r
if(side=="two-sided")
  emp_p_value <- sum(abs(perm_t_values) > t_value_obs)/NumSim

if(side=="one-sided")
  emp_p_value <- sum(perm_t_values > t_value_obs)/NumSim

mean_perm_t_values <- mean(perm_t_values)

#Output of the function
list(t_value_obs = t_value_obs, p_value_obs = p_value_obs,
     mean_perm_t_values = mean_perm_t_values,
     emp_p_value = emp_p_value)
```

Results

Permutation T–test

![Permutation T-test chart]

- Frequency
- Mean permutated T values
- Observed T value

Resampling methods in R 12/14
Results

Observed sample:

- T value: 1.68
- P value: 0.11

Permutation results:

- Mean of all T values: 0.03
- Empirical p value: 0.11
References