HOTEL_FAM: SAS-IML macro for computing association tests (T_{HP}, T_{GP}, T_{Hf}, T_{Gf}, T_{Hm}, and T_{Gm}) for multiple markers in genetic family studies

Michael Knapp[1]
Ruzong Fan[2]
[1]*IMBIE, Bonn*
[2]*NICHD/NIH, Maryland*
Version 0.1

February 14, 2012

1 Overview

This document describes a SAS-IML (SAS Institute Inc. 1999) macro called `hotel_fam` to calculate the test statistics T_{HP}, T_{GP}, T_{Hf}, T_{Gf}, T_{Hm}, and T_{Gm} for the analysis of family-based association studies. Section 2 explains the parameters of this macro. Section 3 presents an example.

In case of questions and/or problems with this program, you can contact us via e-mail (`knapp@uni-bonn.de` and `fanr@mail.nih.gov`).

The theoretical basis for this program is given by Fan et al. (2005).

Please reference Fan et al. (2005) if you use `hotel_fam` in any published work.

2 Syntax

The macro `hotel_fam` is invoked with the command

%`hotel_fam`(*infile, ma_number, ma_names, outfile, perm_number*)

In the following, the five parameters of `hotel_fam` are described:

1. *infile*
   This parameter refers to an existing SAS file, which contains the data to be analysed. Each observation of this SAS file

corresponds to one individual. The following variables must be present in this SAS file:

- `fam`
  Pedigree identifier. `hotel_fam` implicitly assumes that all families are nuclear families (i.e., parents plus a single affected child).

- `ind`
  Number of the individual within a family.

- `father`

- `mother`
  `hotel_fam` only distinguishes two types of individuals: parents and children. For parents, both of these two variables contain the values "0". For children, the variables `father` and `mother` contain values $\neq$ "0".

- `sex`
  Contains the value "1" for a male and "2" for a female.

- `aff_stat`
  Affection status. The value of this variable is "2" for affected individuals. Individuals with any value $\neq$ "2" are considered unaffected.

Additionally, the SAS file *infile* has variables which contain the marker data. For each marker locus, there are two such variables. The names of these variables consist of two parts: (i) an identifier for the locus, and (ii) the suffix _1 and _2 for the variable containing the first and the second allele at this locus.
Example: `D5S17_1` and `D5S17_2`.
Both of the variables corresponding to a marker locus contain the value "0" in case that the individual has not been typed for this marker locus. Otherwise, these variables contain the allele numbers for the individual.
NOTE: The variables `fam`, `ind`, `father`, `mother`, `sex`, `aff` and the variables containing the marker data have to be numeric variables.

2. *ma_number*
   This parameter gives the number of marker loci which has to be included in the analysis.
   Example: *3*

3. *ma_names*
   The value of this parameter is a list of the identifiers of the variables which contain the two alleles for the marker loci to be analyzed. The number of identifiers given in *ma_names*

must coincide with the value of *ma_number*.
Example: *D5S17 D5S91 D5S12.*

4. *outfile*
   This parameter refers to the name of a file which contains the output generated by the macro `hotel_fam`. For more information on the output produced by `hotel_fam`, see the example in the following section.

5. *perm_number*
   The value of this parameter gives the number of replicates for determining the permutation based P value of T_{HP} and T_{GP}. In case that *perm_number*=0, the calculation of permutation based P values is suppressed.

3 Example

The example presented in this section analyses the first three marker loci of the data given in the ASCII file `example.dat`. The SAS statements for the analysis of these data are shown in Figure 1 and are contained in the file `example.sas`. Below, the program

---

```
%include "hotel_fam.sas";

filename ein "example.dat";
data examp;
infile ein;
input fam ind father mother sex aff
      snpa_1 snpa_2 snpb_1 snpb_2 snpc_1 snpc_2 snpd_1 snpd_2
      snpe_1 snpe_2 snpf_1 snpf_2 snpg_1 snpg_2 snph_1 snph_2
      snpi_1 snpi_2 snpj_1 snpj_2 snpk_1 snpk_2 snpl_1 snpl_2
      snpm_1 snpm_2 snpn_1 snpn_2;
run;

%hotel_fam(examp,3,snpa snpb snpc,testdata.out,100000);
```

Figure 1: `example.sas`

`example.sas` is discussed in detail:

```
%include "hotel_fam.sas";
```

This statement includes the file `hotel_fam.sas`, which contains the definition of the macro `hotel_fam`. (It is assumed that the file `hotel_fam` resides in the working directory. Otherwise, an appropriate path has to be provided.)

```
filename ein "example.dat";
data examp;
infile ein;
```

```
input fam ind father mother sex aff
      snpa_1 snpa_2 snpb_1 snpb_2 snpc_1 snpc_2 snpd_1 snpd_2
      snpe_1 snpe_2 snpf_1 snpf_2 snpg_1 snpg_2 snph_1 snph_2
      snpi_1 snpi_2 snpj_1 snpj_2 snpk_1 snpk_2 snpl_1 snpl_2
      snpm_1 snpm_2 snpn_1 snpn_2;
run;
```

This part of the program generates the SAS file examp. Again, it is assumed that the file example.dat resides in the working directory.

```
%hotel_fam(examp,3,snpa snpb snpc,testdata.out);
```

This statement invoke the macro hotel_fam.

Running the program example.sas now generates a file TESTDATA.OUT, which is shown in Figure 2.

```
Marker: snpa snpb snpc

                 no of
control       families coding      statistic  df  p-value
father and mother   114 haplotype      7.94    3 4.72E-02
                        genotype       9.62    6 1.42E-01
father              118 haplotype      3.07    3 3.81E-01
                        genotype       4.39    6 6.24E-01
mother              117 haplotype     11.29    3 1.02E-02
                        genotype      12.05    6 6.08E-02


Permutation based (100000 replicates) p-values for father and mother
haplotype coding:  5.38E-02  genotype coding:  1.00E-01
```

Figure 2: TESTDATA.OUT generated by the SAS program example.sas

---

4 Miscellaneous information

The maximum number of alleles at a single locus is 100. You can increase this maximum number by changing the statement

```
%let max_allele=100
```

in hotel_fam.sas.

The maximum number of families is 1000. You can increase this maximum number by changing the statement

```
%let max_fam=1000
```

in hotel_fam.sas.

5 References

Fan R, Knapp M, Wjst M, Zhao C, Xiong M (2005) High resolution association studies of complex diseases using parents as controls. Annals of Human Genetics 69:187-208.

SAS Institute Inc. (1999) SAS macro language: reference, version 8. SAS Institute, Cary, NC.