

---

# HOTEL\_CC: SAS-IML macro for computing association tests ( $T_G$ and $T_H$ ) for multiple markers in genetic case-control studies

---

Michael Knapp<sup>1</sup>

Ruzong Fan<sup>2</sup>

<sup>1</sup>IMBIE, Bonn

<sup>2</sup>NICHD/NIH, Maryland

Version 0.1

February 14, 2012

## 1 Overview

This document describes a SAS-IML (SAS Institute Inc. 1999) macro called `hotel_cc` to calculate the test statistics  $T_G$  and  $T_H$  for the analysis of case-control 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 and Knapp (2003).

Please reference Fan and Knapp (2003) if you use `hotel_cc` in any published work.

## 2 Syntax

The macro `hotel_cc` is invoked with the command

`%hotel_cc(infile, aff, ma_number, ma_names, outfile)`

In the following, the five parameters of `hotel_cc` 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 SAS file *infile* must contain a variable `<affection status>`, which describes the affection status of the individual. The name of the variable

which contains the affection status is the second parameter of `hotel_cc` (see below). 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 variable `<affection status>` and the variables containing the marker data have to be numeric variables.

2. *aff*

This parameter is the name of the variable of *infile* which contains the affection status of the individual. A value of 2 for this variable indicates an affected individual, whereas all other values indicate an unaffected individual.

3. *ma\_number*

This parameter gives the number of marker loci which has to be included in the analysis.

Example: 3

4. *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`.

5. *outfile*

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

### 3 Example

The example presented in this section analyses the two 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 `example.sas` is discussed in detail:

```
%include "hotel_cc.sas";
```

This statement includes the file `hotel_cc.sas`, which contains the definition of the macro `hotel_cc`. (It is assumed that the file

---

```

%include "hotel_cc.sas";

filename ein "example.dat";
data examp;
infile ein;
input aff_stat first_1 first_2 second_1 second_2;
run;

%hotel_cc(examp,aff_stat,2,first second,RESULT.OUT);

```

Figure 1: example.sas

---

`hotel_cc` resides in the working directory. Otherwise, an appropriate path has to be provided.)

```

filename ein "example.dat";
data examp;
infile ein;
input aff_stat first_1 first_2 second_1 second_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_cc(examp,aff_stat,2,first second,RESULT.OUT);

```

This statement invoke the macro `hotel_cc`.

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

---

Marker: first second

coding	statistic	df	p-value
Haplotype	6.98	4	1.37E-01
Genotype	12.75	11	3.10E-01

Figure 2: `RESULT.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_cc.sas`.

## 5 References

Fan R, Knapp M (2003) Genome association studies of complex diseases by case-control designs. *Am J Hum Genet* 72: 850–868.

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