Experimental Designs for Sensory Trials:

Abstract rules and practical requirements

Joachim Kunert, TU Dortmund

With the design of sensory experiments, sensometricians (including referees of journals) appear to have only one attitude:

use a design which is as balanced as possible...

(whatever might be the purpose of the experiment).

In this talk,

I try to show why the recommendation

"Use a design which is as balanced as possible"

is not always a good idea.

<u>Randomized versus systematic design</u>

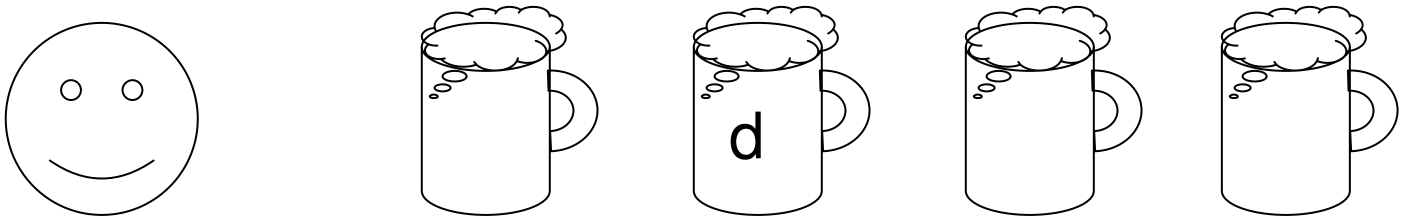
We consider a very general model

$y = \tau + \phi + e$

with

y	....	real observation
τ	....	ideal observation
ϕ	....	systematic bias due to nuisance
e	....	random error

Nuisance ϕ



$$y_{ij} = \tau_{d(i,j)} + \phi_{ij} + e_{ij}$$

One approach of experimental design would be to

- find a detailed model for the nuisance ϕ

- assume that the errors are iid.

Once we have a model for ϕ, we then can determine an experimental design which is in some sense optimal for this model.

This normally leads to designs with a high degree of balance.

Modelling ϕ.

A popular example is the row-column model:

$$y_{ij} = \tau_{d(i,j)} + \mu + \alpha_i + \beta_j + e_{ij}$$

Here, ϕ is modelled as

an effect of the assessor + an effect of the period.

For a simulation experiment, we look at the simple case that we have

4 assessors

4 products

and each assessor can evaluate 4 products.

If we assume that the row- column model is correct, we get ϕ_{ij} with a restricted structure, like

$$\phi = \begin{bmatrix} 2 & 3 & 4 & 5 \\ 7 & 8 & 9 & 10 \\ 4 & 5 & 6 & 7 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

Use this as the basis of a simulation experiment.

A good design for the row-column model is

a Latin square, like

$$d_1 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \\ 3 & 1 & 4 & 2 \\ 4 & 3 & 2 & 1 \end{bmatrix}$$

Using ϕ and adding a random error, we construct data of the following structure

$$y = \begin{bmatrix} y_{11} & y_{12} & y_{13} & y_{14} \\ y_{21} & y_{22} & y_{23} & y_{24} \\ y_{31} & y_{32} & y_{33} & y_{34} \\ y_{41} & y_{42} & y_{43} & y_{44} \end{bmatrix}$$

Analyzing these data, assuming design d_1, we get estimates like

$$\widehat{\tau_1 - \tau_3} = \frac{1}{4}(y_{11} + y_{23} + y_{32} + y_{44}) - \frac{1}{4}(y_{13} + y_{24} + y_{31} + y_{42})$$

The row- and the column-effect cancel in this estimate

Neglecting the period effect might lead to a design like

$$d_2 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

Here we would get the estimate

$$\widehat{\tau_1 - \tau_3} = \frac{1}{4}(y_{11} + y_{21} + y_{31} + y_{41}) - \frac{1}{4}(y_{13} + y_{23} + y_{33} + y_{43})$$

where the period effect does not cancel.

Simulating 1,000 data sets from the row- column model and calculating $\widehat{\tau_1 - \tau_3}$, assuming designs d_1 and d_2, we get

design	mean	standard deviation
systematic latin square d_1	0.009	0.711
complete block design d_2	-2.006	0.706

For systematic designs,
neglecting an effect which is present in the data
leads to biassed estimates

The alternative would be a randomized design.

Randomizing the order for each assessor independently might lead to

$$d_3 = \begin{bmatrix} 2 & 4 & 1 & 3 \\ 4 & 1 & 2 & 3 \\ 3 & 1 & 2 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

Here we would get the estimate

$$\widehat{\tau_1 - \tau_3} = \frac{1}{4}(y_{13} + y_{22} + y_{32} + y_{41}) - \frac{1}{4}(y_{14} + y_{24} + y_{31} + y_{43})$$

where the period effect still does not cancel.

Randomization does not imply
that we get a balanced design.

Randomization theory considers the observed design as one possible outcome of an experiment,
just as the observed data are also one outcome among many possibilities.

Simulating 1,000 data sets from the row- column model and calculating $\widehat{\tau_1 - \tau_3}$, we get

design	mean	standard deviation
systematic latin square d_1	0.009	0.711
randomized complete block design	-0.015	1.16

For randomized designs,
neglecting an effect which is present in the data
leads to an increased variance
(but leaves the estimates unbiassed)

It is possible to use randomization and consider the period effects:

A randomized latin square
randomizes rows and columns (as a whole)

Simulating 1,000 data sets from the row- column model and calculating $\widehat{\tau_1 - \tau_3}$, we get

design	mean	standard deviation
systematic latin square d_1	0.009	0.711
randomized complete block design	-0.015	1.16
randomized latin square	0.005	0.718

The advantage of randomization gets visible if the model is not exactly correct.

Assume

$$\phi = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 9 & 12 \\ 4 & 8 & 12 & 16 \end{bmatrix}$$

There is a row- column-structure, but not exactly additive.

Simulating 1,000 data sets from this nonlinear ϕ, and calculating $\widehat{\tau_1 - \tau_3}$, we get

design	mean	standard deviation
systematic latin square d_1	1.756	0.706
randomized complete block design	0.006	2.58
randomized latin square	0.011	1.84

The systematic latin square is no longer unbiassed.

Both kinds of randomized designs remain unbiassed, the variance of the estimates gets better for a randomization that is adapted to the structure of the data.

Note that even the randomized design
which neglects the period effect present in the data
performs better than the systematic latin square

Carryover effects are a special problem.

If additive carryover effects are present in the data,

than neighbour balanced designs minimize

(a function of) the bias due to carryover effects.

Looking through FQaP, virtually all
articles which discuss their design appear to stress that

"the products were accorded to assessors according to a
Williams latin square, balanced for carryover effects".

(I would have been a lot happier to read something about
washout periods or other methods to avoid carryover.)

There are three major problems:

1) Even with balanced designs carryover effects are not orthogonal to direct effects, hence uncorrected estimates are not unbiassed.

2) For the corrected estimates, there is no justification of iid errors.

3) The model with additive carryover effects is at best approximately valid.

Further problem with carryover:

Do we even want to estimate the direct effect?

What about permanent effects?

(permanent effect = direct effect + carryover)

If we want consumers to taste more than a small amount of our product, they will experience the "permanent effect".

For the estimation of permanent effects,

neighbour balanced latin squares are not efficient!

Optimal design for permanent effects

(Bailey and Druilhet, 2004)

$$d_4 = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 3 & 3 \\ \vdots & & & \vdots \\ 4 & 4 & 3 & 3 \end{bmatrix}$$

Period effects:

Even balancing for period effects is not always a good idea.

Assume we do not want to compare the products, but want to find out whether there are differences between groups of assessors: Do assessors 1 and 2 have a better preference for product 1 than assessors 3 and 4?

Then we would simply give product 1 to all four assessors.

Now assume we want to repeat this for products 2, 3 and 4.

With the latin square

$$d_1 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \\ 3 & 1 & 4 & 2 \\ 4 & 3 & 2 & 1 \end{bmatrix}$$

the measurement of product 1

by assessor 1 contains period 1,

by assessor 2 contains period 3,

by assessor 3 contains period 2,

by assessor 4 contains period 4.

With the design

$$d_2 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

all measurements of product 1 contain period 1,

and the difference between the assessors is unbiassed by period effects.

We used the same idea of a design
in an experiment,
where we compared the food preferences and sensitivity of
obese, over-weight and normal weight
children for the taste categories,

like sweet, fat, sour, etc.

It took us some time until we found a journal with referees who appreciated the design.

References:
Alexy et al. (2011). *Journal of Sensory Studies*.
Bailey & Druilhet (2004). *Annals of Statistics*.
Kunert (1998). *Food Quality and Preference*.