7th Sensometrics Meeting
July 28 – 30 2004, UC Davis

On nearly balanced designs for sensory trials

Oliver Sailer
Joachim Kunert

SFB 475 – Reduction of Complexity for Multivariate Data Structures
Department of Statistics
University of Dortmund
Dortmund, Germany

Outline

1. Introduction
2. Validity and the row-column model
3. Simulation study
    - Validity of balanced designs
    - Performance of nearly balanced designs
    - Performance of imbalanced designs
4. Summary

1. Introduction

- We want to design a sensory experiment for the comparison of several products.
- Each assessor tastes some / all of the products.

Product	Assessor 1	Assessor 2	Assessor 3	…
Period 1	2	1	5	…
Period 2	5	3	4	…
Period 3	3	2	1	…
…	…	…	…	…

- Possible Problems to deal with
    - Period effects
    - Carryover effects
    - Correlation of observations due to other effects
- Possible solutions
    - Use balanced designs
        - balanced for subject effects
        - balanced for period effects / uniform on the periods
        - balanced for (first order) carryover effects
    - Randomize
- Can we combine these solutions in the row-column model?

2. Validity and the row-column model

Balanced experimental designs

- There are t products, b asseessors and $k \leq t$ periods
- Generalized Youden design (GYD)

  Each assessor receives each product at most once.

  The number of assessors receiving any pair of distinct treatments is a constant.

  Each product appears in each period either $\left[\frac{b}{t}\right]$ or $\left[\frac{b}{t}\right]+1$ times.

  The number of periods where any pair of distinct treatments occurs is a constant.
- Uniform on the periods

  Each product appears in each period equally often.
- Carryover balance

  Each product is preceded by every other product equally often.

The row-column model

- $Y = T_d\tau + (1_b \otimes I_k)\alpha + (I_b \otimes 1_k)\beta + e$

  where Y is the response vector, T_d is the treatment design matrix, τ are the direct effects of the products, α the period effects, β the effects of the assessors and $e \sim (0, \sigma^2 I_{bk})$ is the error term.

- Keep the responses fixed and randomize w.r.t the design.

- valid model (Bailey and Rowley, 1987): The row-column model is valid for a given randomization if for any contrast $\bullet\tau$

$$E_\delta(\bullet\hat{\tau}) = \bullet\tau$$

and

$$E_\delta[\hat{\text{var}}(\bullet\hat{\tau})] = \text{var}_\delta(\bullet\hat{\tau}).$$

Theorem 1

- If we randomize assessors and product orders, balance for carryover effects is destroyed.

- Idea based on Bailey (1985)

  Start with a generalized Youden design that is uniform on the periods with desirable carryover structure.

  Randomize product labels.

  Randomize assessors.

- Theorem 1 (Kunert, 1998): If there are no carryover effects, this procedure validates the row-column model and preserves the carryover structure.

- Some authors suggest using “nearly“ balanced designs
  - Ball (1997)
  - Périnel and Pagès (2004)
- What happens if
  - we use GYDs that aren‘t uniform on the periods?
  - we don't use generalized Youden designs?
  - we don‘t use carryover balanced designs?
- Assess the validity of the model when
  - there are no carryover effects.
  - carryover effects are present.

Theorem 2

- Consider a GYD where each assessor gets each product once and the number of assessors is not a multiple of the number of products.
- Randomize assessors and product labels.
- Assume that there are no carryover effects.
- Then this randomization still validates the row-column model.

Theorem 3

- Consider a generalized Youden design.
- Randomize assessors and product labels.
- Assume that there are carryover effects ρ.
- Then the contrast estimates from the row-column model will generally be biased.

- Theorem 3: If the starting design and the design which consists of the first $k-1$ periods of it are both GYDs that are uniform on the periods and if the starting design is balanced for carryover effects, then the variance of the estimate is the same with or without carryover effects.

3. Simulation Study

- Use artificial data with period effects.
- Randomize 10000 designs.
- For each permutation i compute $\bullet\hat{\tau}^{(i)} = \hat{\tau}_1 - \hat{\tau}_t^{(i)}$ and $\hat{\mathrm{var}}(\bullet\hat{\tau})^{(i)}$.
- Use these statistics to assess the validity of the row-column model.

- Assess unbiasedness of the contrast estimate

  Tests / confidence intervals on arithmetic means of contrast estimates based on central limit theorem.

  Only of interest when there are carryover effects.

- Assess unbiasedness of the variance estimate of the contrast estimate

  Partition the permutations in 100 groups.

  Empirical variance of the contrast in each group is a good estimate of the true variance.

  Arithmetic mean of the variance estimates from the linear model is a good estimate of the mean of the model-based variance estimates.

  Tests / confidence intervals on difference of the two estimates based on central limit theorem.

- Compute permutation t-statistics $t^{(i)} = \frac{\bullet\hat{\tau}^{(i)}}{\sqrt{\hat{\mathrm{var}}(\bullet\hat{\tau})^{(i)}}}$.

  Compare EDF of the t-statistics with the CDF of the t-distribution with $(b-1)(k-1)-t+1$ df.

  Compute the EDF at the 5%-quantile of the t-distribution.

Validity of carryover balanced GYDs

- Example (Kunert, 1998)

$$\begin{pmatrix} 2&1&5&4&3&5&4&3&2&1 \\ 3&2&1&5&4&4&3&2&1&5 \\ 1&5&4&3&2&1&5&4&3&2 \\ 4&3&2&1&5&3&2&1&5&4 \\ 5&4&3&2&1&2&1&5&4&3 \end{pmatrix}$$

- This design is composed of two latin squares.

  It is a GYD.

  It is uniform on the periods.

- This design is carryover balanced.
- The first 9 columns of the design still form a GYD uniform on the periods.
- Theorems 1 and 3 above hold.

- Results for 10000 permutations of the design

	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{32,0.05}$	Mean of $\bullet\hat{\tau}^{(i)}$	$\bullet\tau$	Emp. var. of $\bullet\hat{\tau}^{(i)}$	Mean of $\hat{\text{var}}(\bullet\hat{\tau})^{(i)}$
Case 1	0	0	0	0.0505	-0.0109	0	0.2482	0.2486
Case 2	-1	0	0	0.6238	-1.0109	-1	0.2482	0.2486
Case 3	0	0	10	0.0000	-0.0109	0	0.2482	3.8502
Case 4	-1	0	10	0.0000	-1.0109	-1	0.2482	3.8502
Case 5	0	10	0	0.0059	-2.0109	0	0.2482	3.8490
Case 6	-1	10	0	0.3114	-3.0109	-1	0.2482	3.8490

- Case 1: No direct and carryover effects.
  Estimate of the variance is unbiased: 95%-CI (-0.0059,0.0058) includes 0
- Case 2: Direct effect of product 1.
  Contrast estimaste $\bullet\hat{\tau} = \hat{\tau}_1 - \hat{\tau}_5$ is shifted by -1, variance stays the same.

- Results for 10000 permutations of the design

	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{32,0.05}$	Mean of $\bullet\hat{\tau}^{(i)}$	$\bullet\tau$	Emp. var. of $\bullet\hat{\tau}^{(i)}$	Mean of $\hat{\mathrm{var}}(\bullet\hat{\tau})^{(i)}$
Case 1	0	0	0	0.0505	-0.0109	0	0.2482	0.2486
Case 2	-1	0	0	0.6238	-1.0109	-1	0.2482	0.2486
Case 3	0	0	10	0.0000	-0.0109	0	0.2482	3.8502
Case 4	-1	0	10	0.0000	-1.0109	-1	0.2482	3.8502
Case 5	0	10	0	0.0059	-2.0109	0	0.2482	3.8490
Case 6	-1	10	0	0.3114	-3.0109	-1	0.2482	3.8490

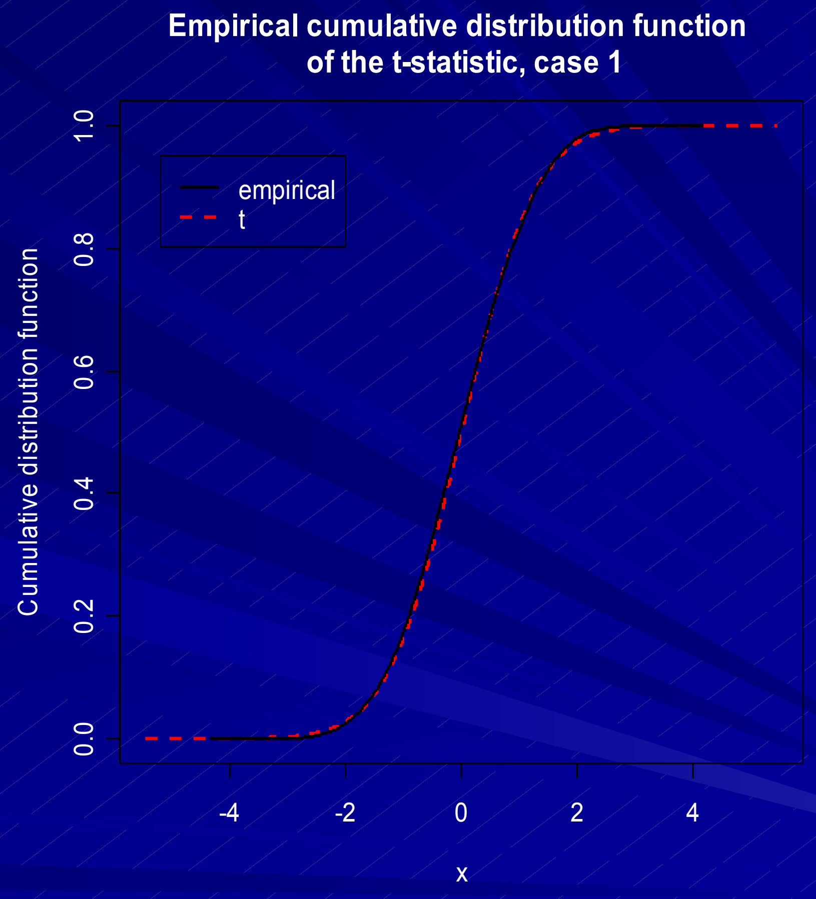
- Case 3: Carryover effect of product 2
  True variance does not increase.
  Estimate of the variance is inflated by factor 15.
- Case 4:
  Contrast estimaste shifted by -1, variance stays the same.

- Results for 10000 permutations of the design

	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{32,0.05}$	Mean of ●$\hat{\tau}^{(i)}$	●τ	Emp. var. of ●$\hat{\tau}^{(i)}$	Mean of $\hat{\text{var}}$(●$\hat{\tau})^{(i)}$
Case 1	0	0	0	0.0505	-0.0109	0	0.2482	0.2486
Case 2	-1	0	0	0.6238	-1.0109	-1	0.2482	0.2486
Case 3	0	0	10	0.0000	-0.0109	0	0.2482	3.8502
Case 4	-1	0	10	0.0000	-1.0109	-1	0.2482	3.8502
Case 5	0	10	0	0.0059	-2.0109	0	0.2482	3.8490
Case 6	-1	10	0	0.3114	-3.0109	-1	0.2482	3.8490

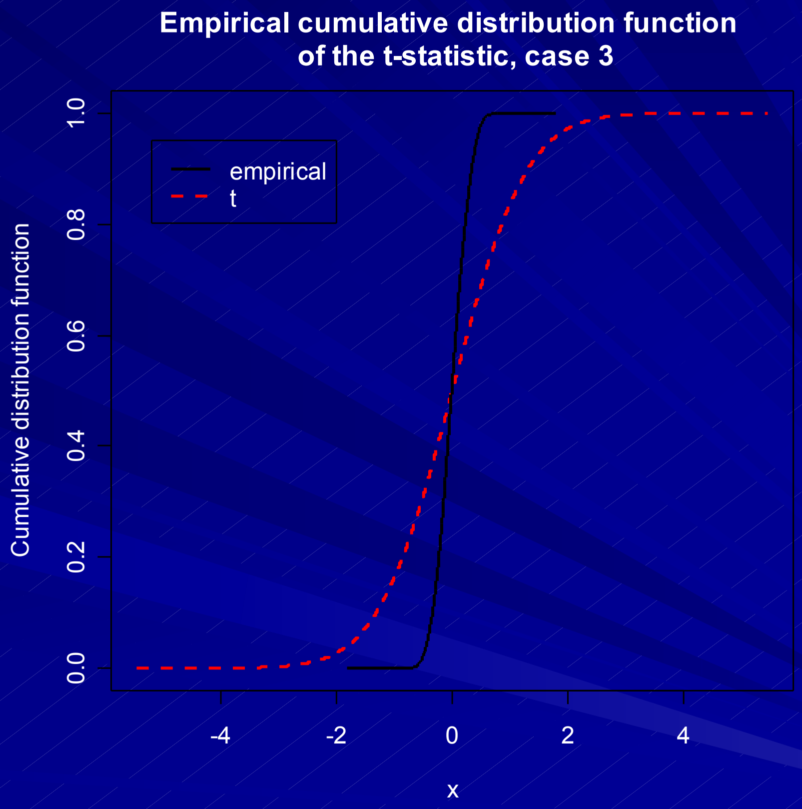
- Case 5: Carryover effect of product 1
  Estimate of the constrast is biased.
  Variance does not increase.
  Estimate of the variance is inflated by factor 15.
- Case 6:
  Contrast estimaste shifted by -1, variance stays the same.

- CDF of the permutation t-statistics and CDF of the t-distribution in case 1
- Perfect fit.
- Again demonstrates that row-column model is valid.



- CDF of the permutation t-statistics and CDF of the t-distribution in case 3
- EDF has steep slope:

  Variance is overestimated.
- t-statistics are too small.
- Conservative analysis



- Let's take a look at the design again

$$\begin{pmatrix} 2\,1\,5\,4\,3 & 5\,4\,3\,2\,1 \\ 3\,2\,1\,5\,4 & 4\,3\,2\,1\,5 \\ 1\,5\,4\,3\,2 & 1\,5\,4\,3\,2 \\ 4\,3\,2\,1\,5 & 3\,2\,1\,5\,4 \\ 5\,4\,3\,2\,1 & 2\,1\,5\,4\,3 \end{pmatrix}$$

- Let's take a look at the design again

$$\begin{pmatrix} 2\ 1\ 5\ 4\ 3 & 5\ 4\ 3\ 2\ 1 \\ 3\ 2\ 1\ 5\ 4 & 4\ 3\ 2\ 1\ 5 \\ 1\ 5\ 4\ 3\ 2 & 1\ 5\ 4\ 3\ 2 \\ 4\ 3\ 2\ 1\ 5 & 3\ 2\ 1\ 5\ 4 \\ 5\ 4\ 3\ 2\ 1 & 2\ 1\ 5\ 4\ 3 \end{pmatrix}$$

- Delete the last assessor

- Let's take a look at the design again

$$\begin{pmatrix} 2\,1\,5\,4\,3 & 5\,4\,3\,2 \\ 3\,2\,1\,5\,4 & 4\,3\,2\,1 \\ 1\,5\,4\,3\,2 & 1\,5\,4\,3 \\ 4\,3\,2\,1\,5 & 3\,2\,1\,5 \\ 5\,4\,3\,2\,1 & 2\,1\,5\,4 \end{pmatrix}$$

- The resulting design still forms a GYD uniform on the periods.
- Theorem 2 holds.
- The design is only nearly carryover balanced.
- The last 8 periods of the design no longer form a GYD.

- Results for 10000 permutations of the design

	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{28,0.05}$	Mean of $\bullet\hat{\tau}^{(i)}$	$\bullet\tau$	Emp. var. Of $\bullet\hat{\tau}^{(i)}$	Mean of $\hat{\text{var}}(\bullet\hat{\tau})^{(i)}$
Case 1	0	0	0	0.0488	0.0109	0	0.3132	0.3103
Case 2	-1	0	0	0.5233	-0.9891	-1	0.3132	0.3103
Case 3	0	0	10	0.0000	0.0064	0	0.9075	4.3542
Case 4	-1	0	10	0.0029	-0.9936	-1	0.9075	4.3542
Case 5	0	10	0	0.0195	-1.9811	0	0.5656	4.3618
Case 6	-1	10	0	0.2446	-2.9811	-1	0.5656	4.3618

- Case 1: Estimate of the variance is unbiased: 95%-CI (-0.0058,0.0094) includes 0.
- Case 3: The variance increases, but is still overestimated: conservative analysis.
- Case 5: The contrast estimate is biased. The variance is higher than in case 1, but is still overestimated.

Performance of nearly carryover balanced GYDs

- A design proposed by Périnel and Pagès (2004)

$$\begin{pmatrix} 7 & 5 & 10 & 7 & 6 & 4 & 1 & 8 & 6 & 5 & 9 & 2 & 3 & 9 & 1 & 8 & 10 & 2 & 3 & 4 & 6 & 3 & 7 & 9 & 8 & 1 & 5 & 10 & 4 & 2 \\ 8 & 3 & 9 & 4 & 1 & 3 & 5 & 4 & 9 & 6 & 6 & 8 & 2 & 10 & 7 & 2 & 1 & 10 & 5 & 7 & 4 & 10 & 3 & 8 & 9 & 6 & 2 & 7 & 5 & 1 \\ 10 & 6 & 4 & 10 & 8 & 7 & 4 & 1 & 2 & 8 & 3 & 3 & 9 & 7 & 2 & 6 & 5 & 5 & 1 & 9 & 2 & 8 & 10 & 1 & 3 & 5 & 4 & 6 & 7 & 9 \\ 5 & 7 & 8 & 2 & 2 & 6 & 3 & 10 & 4 & 9 & 1 & 9 & 8 & 1 & 3 & 10 & 7 & 4 & 6 & 5 & 9 & 5 & 4 & 2 & 7 & 10 & 8 & 1 & 10 & 6 \\ 2 & 10 & 6 & 1 & 7 & 5 & 8 & 6 & 5 & 1 & 4 & 10 & 7 & 3 & 4 & 3 & 9 & 9 & 2 & 8 & 7 & 1 & 6 & 5 & 4 & 2 & 3 & 9 & 8 & 3 \end{pmatrix}$$

- This design is not a GYD, not even a balanced incomplete block design.
- It is not carryover balanced.
- But only small departures from ideal structure.

  The design is “nearly” balanced.

- Results for 10000 permutations of the design

	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{107,0.05}$	Mean of $\bullet\hat{\tau}^{(i)}$	$\bullet\tau$	Emp. var. Of $\bullet\hat{\tau}^{(i)}$	Mean of $\hat{\text{var}}(\bullet\hat{\tau})^{(i)}$
Case 1	0	0	0	0.0507	-0.0017	0	0.1708	0.1684
Case 2	-1	0	0	0.7807	-0.9983	-1	0.1708	0.1684
Case 3	0	0	10	0.0024	0.0047	0	0.4731	1.3932
Case 4	-1	0	10	0.0847	-0.9953	-1	0.4731	1.3932
Case 5	0	10	0	0.5221	-1.9922	0	0.3169	1.3932
Case 6	-1	10	0	0.9647	-2.9922	-1	0.3169	1.3932

- Case 1:

  Estimate of the variance of $\bullet\hat{\tau} = \hat{\tau}_1 - \hat{\tau}_5$ seems to be unbiased: 95%-CI (-0.0021,0.0068) includes 0.
  If the design is not valid, the bias is very small.

- Results for 10000 permutations of the design

	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{107,0.05}$	Mean of $\bullet\hat{\tau}^{(i)}$	$\bullet\tau$	Emp. var. Of $\bullet\hat{\tau}^{(i)}$	Mean of $\hat{\text{var}}(\bullet\hat{\tau})^{(i)}$
Case 1	0	0	0	0.0507	-0.0017	0	0.1708	0.1684
Case 2	-1	0	0	0.7807	-0.9983	-1	0.1708	0.1684
Case 3	0	0	10	0.0024	0.0047	0	0.4731	1.3932
Case 4	-1	0	10	0.0847	-0.9953	-1	0.4731	1.3932
Case 5	0	10	0	0.5221	-1.9922	0	0.3169	1.3932
Case 6	-1	10	0	0.9647	-2.9922	-1	0.3169	1.3932

- Case 3: The variance increases, but is still overestimated: conservative analysis.
- Case 5: The contrast estimate is biased. The variance is higher than in case 1, but is still overestimated.

Performance of a strongly imbalanced design

- So far we only looked at designs with a “small degree of imbalance“.

$$\begin{pmatrix} 1&1&1&1&1&1&1&3&2&2&2&2&4&5&4 \\ 2&2&2&2&3&3&3&1&3&3&3&3&5&4&5 \\ 4&4&5&5&4&4&5&5&4&4&5&5&1&2&3 \end{pmatrix}$$

- This design is no GYD, not even a balanced incomplete block design.
- It is also not balanced for carryover effects.

- Results for 10000 permutations of the design

	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{24,0.05}$	Mean of $\bullet\hat{\tau}^{(i)}$	$\bullet\tau$	Emp. var. of $\bullet\hat{\tau}^{(i)}$	Mean of $\hat{\text{var}}(\bullet\hat{\tau})^{(i)}$
Case 1	0	0	0	0.0492	0.0121	0	0.7084	0.6699
Case 2	-1	0	0	0.3104	-0.9879	-1	0.7084	0.6699
Case 3	0	0	10	0.0890	0.0072	0	5.6120	3.4831
Case 4	-1	0	10	0.2054	-0.9928	-1	5.6120	3.4831
Case 5	0	10	0	0.5078	-3.3571	0	3.8175	3.4567
Case 6	-1	10	0	0.7324	-4.3571	-1	3.8175	3.4567

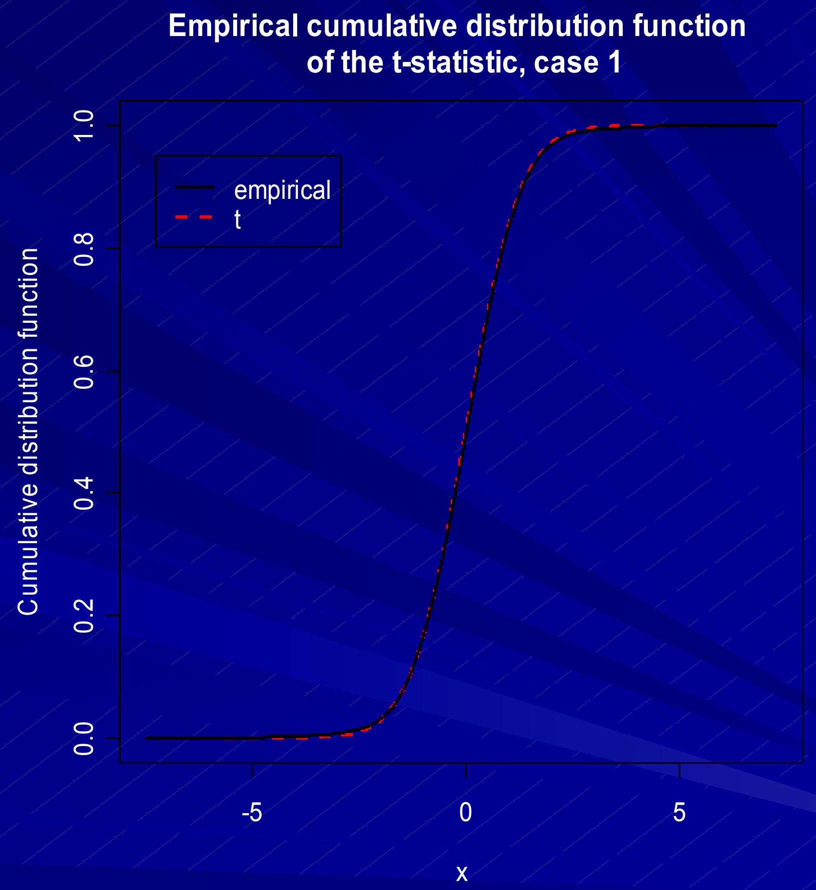
- Case 1:
  Estimate of the variance is biased: 95%-CI (0.0164,0.0609) does not include 0.
  Is this bias relevant?

- Results for 10000 permutations of the design

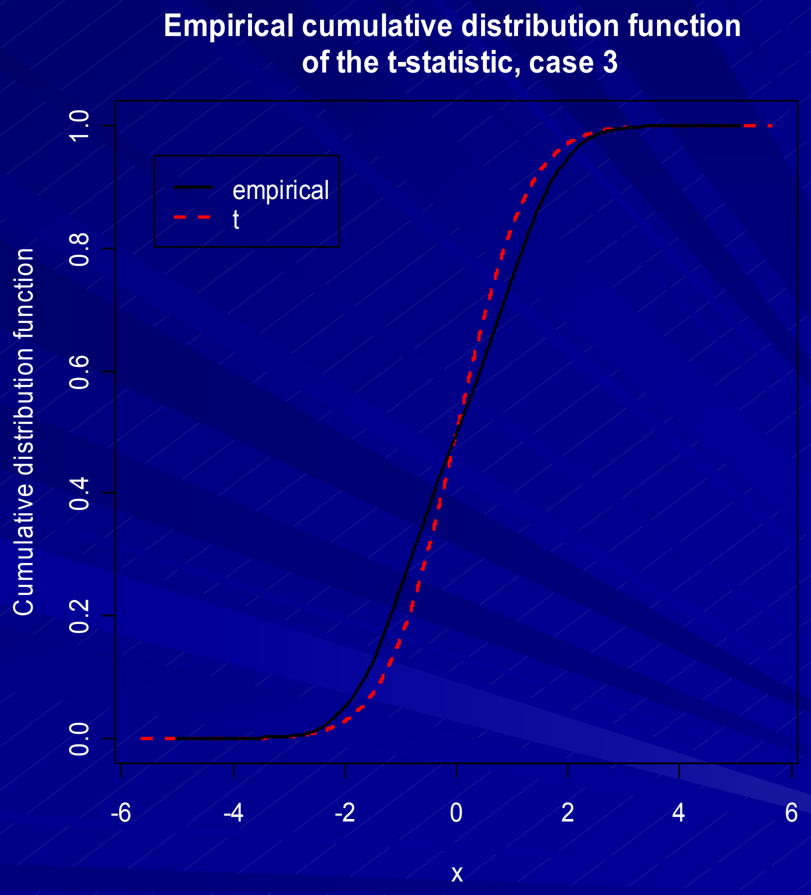
	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{24,0.05}$	Mean of $\bullet\hat{\tau}^{(i)}$	$\bullet\tau$	Emp. var. of $\bullet\hat{\tau}^{(i)}$	Mean of $\hat{\text{var}}(\bullet\hat{\tau})^{(i)}$
Case 1	0	0	0	0.0492	0.0121	0	0.7084	0.6699
Case 2	-1	0	0	0.3104	-0.9879	-1	0.7084	0.6699
Case 3	0	0	10	0.0890	0.0072	0	5.6120	3.4831
Case 4	-1	0	10	0.2054	-0.9928	-1	5.6120	3.4831
Case 5	0	10	0	0.5078	-3.3571	0	3.8175	3.4567
Case 6	-1	10	0	0.7324	-4.3571	-1	3.8175	3.4567

- Case 3: The variance is inflated by a factor of 8.
  The variance estimate clearly is too small.
  Anti-conservative analysis (empirical level of the t-test is 8.9%)
- Case 5:
  The contrast estimate is biased. The variance is underestimated.

- CDF of the permutation t-statistics and CDF of the t-distribution in case 1
- Design is clearly imbalanced.
- But: t-distribution still seems to fit well.



- CDF of the permutation t-statistics and CDF of the t-distribution in case 3
- Variance is underestimated.
- t-test is no longer conservative.



4. Summary

- Randomization of treatment labels and assessors validates the row-column model (when there are no carryover effects) for special types of balanced GYDs.
- If the various types of balance are only slightly violated, the analysis
  - may still be valid.
  - even if it is invalid, it is not changed much.
- If there are carryover effects, the estimates are biased whether we use carryover balanced designs or not.
- Slight violations still don't make much of a difference.
- Departures from carryover balance seem to be more serious than departures from the GYD conditions.

References

- Bailey, R.A. (1985). Restricted randomization for neighbour-balanced designs, Statistics and Decisions Supplement 2, 214-223.
- Bailey, R. and Rowley, C.A. (1987). Valid randomization, Proceedings of the Royal Society London A 410, 105-124.
- Ball R. D. (1997): Incomplete block designs for the minimisation of order and carry-over effects in sensory analysis, Food Quality and Preference 8(2), 111-118.
- Kunert, J. (1998). Sensory experiments as crossover studies, Food Quality and Preference 9, 243-253.
- Pèrinel, E. and Pagès, J. (2004): Optimal nested crossover designs in sensory analysis, Food Quality and Preference 15, 439-446.