Smoothing the interaction of two categorical predictors

(This is not in RWC.)

I'll present balanced ANOVA with one error term, as in

Hodges, Cui, Sargent, Carlin (*Technometrics* 2007).

Cui & Hodges (manuscript; web site) do the general balanced case.

We call this "Smoothed ANOVA" and our analysis uses MLMs.

Other versions of smoothed ANOVA:

Gelman (Ann Stat 2005) also uses MLMs.

Nobile & Green (Biometrika 2000) doesn't use MLMs.

Smoothed ANOVA, general and specific motivation

General motivation: Statistical folklore and experience say that

- interactions are often absent or small, but
- it's unwise to assume any specific interaction is absent.

Many analyses use significance tests to delete interactions, but
model-averaging and smoothing outperform stepwise methods.

This suggests shrinking or smoothing interactions instead.

Specific motivation: Dixon & Simon's (1991) subgroup analysis.

Treatment effects in (say) males and females are considered after first shrinking the treatment-by-sex interaction.

This filters out some error and shrinks spurious apparent differences between subgroups.

Dataset: Soft denture-liner materials

Soft denture liners are fabricated on a hard denture base, then polished and finished.

Polishing and finishing can leave a gap between the liner and base.

Such gaps harbor *Candida* and other oral pathogens, which is bad.

Pesun et al. (2002) compared gaps, measured in μm, for

- 2 soft-liner materials, standard and new (factor M), with
- 4 four polishing methods (factor P) and
- 8 finishing methods (factor F), with
- no replication within design cells.

We analyzed $\log_{10}$gap.

Problems, which smoothed ANOVA addresses

(1) Standard analysis: Use the highest-order interaction as the error term.

But there's an outlier and it matters with this approach.

Analysis	M × P	M × F	P × F
All data	0.12	0.097	0.15
Omit outlier	0.096	0.004	0.16

Are there alternatives to this error term and a keep/omit choice for the outlier?

(2) The P×F and M×P×F interactions have 21 DF each. The (likely) many null contrasts dilute the (likely) few "live" ones.

(3) Special interest: M×P and M×F, with 3 and 7 DF. Shrinking (smoothing) these would reduce clutter.

Model set-up and notation

I'll develop this using a 2^3 design with 6 replicates per design cell.

Assume: balanced design with c cells and $m \geq 1$ reps per cell, so $n = cm$.

2^3 design with 6 reps per cell, $c = 2^3 = 8, m = 6,\ n = cm = 48$.

Write the ANOVA as a linear model with each effect having design-matrix columns orthogonal to each other and to columns for other effects.

$\underline{\mathbf{X}}$: The p columns for the grand mean and main effects, $cm \times p$.

$\underline{\mathbf{Z}}$: The q columns for interactions, $cm \times q$.

Scale the columns of $\mathbf{X}$ and $\mathbf{Z}$ so $\mathbf{X}'\mathbf{X} = cm\mathbf{I}_p$ and $\mathbf{Z}'\mathbf{Z} = cm\mathbf{I}_q$.

Model set-up and notation, continued

The usual ANOVA model is $\mathbf{y} = [\mathbf{X}|\mathbf{Z}]\Theta + \epsilon$, where

- $\mathbf{y}$ is the *cm*-vector of data,
- $\Theta = (\Theta_1', \Theta_2')'$ conforming to $[\mathbf{X}|\mathbf{Z}]$,
- $\boldsymbol{\epsilon} \sim N_{cm}(\mathbf{0}, \mathbf{R})$ for $\mathbf{R} = \frac{1}{\eta_0}\mathbf{I}_{cm}$ with η_0 unknown.

2^3 example: One choice of $[\mathbf{X}|\mathbf{Z}]$ is $H \otimes \mathbf{1}_6$ for $H =$

$$\left[\begin{array}{cccc|cccc}
+1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\
+1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\
+1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\
+1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\
+1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\
+1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\
+1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\
+1 & -1 & -1 & -1 & +1 & +1 & +1 & -1
\end{array}\right].$$

H's first 4 columns give $\mathbf{X}$; its last 4 columns give $\mathbf{Z}$.

The key to smoothed ANOVA: The prior on Θ

The usual ANOVA model is $\mathbf{y} = [\mathbf{X}|\mathbf{Z}]\Theta + \epsilon$, where

$\Theta = (\Theta_1'|\Theta_2')'$ conforming to [main effects | interactions].

The interactions are shrunk by means of $\mathbf{G} = \text{cov}(\Theta_2)$.

In this lecture, for θ_k in $\Theta_2 = (\theta_{p+1}, \ldots, \theta_{p+q})$,

$\theta_k|\phi_k \sim N(0, 1/\phi_k)$ independently given the ϕ_k.

You specify a set of distinct random-effect precisions $\{\eta_1, \ldots, \eta_s\}, s \leq q$, and a function $j(k) \ni \phi_k = \eta_{j(k)}$, so $\mathbf{G}^{-1} = \text{diag}(\eta_{j(1)}, \ldots, \eta_{j(q)})$.

This groups the θ_k and the columns in $\mathbf{Z}$:

group j has n_j θ_k, which are smoothed by η_j.

Let $\eta = (\eta_0, \eta_1, \ldots, \eta_s)$, the error precision and smoothing precisions.

Smoothed ANOVA in mixed-linear-model terms

MLM standard form: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$

SANOVA: $\mathbf{y} = \mathbf{X}\Theta_1 + \mathbf{Z}\Theta_2 + \boldsymbol{\epsilon}$, with $\mathbf{R} = \frac{1}{\eta_0}\mathbf{I}_{cm}$

$$\text{and } \mathbf{G} = \text{cov}(\Theta_2) = \begin{bmatrix} \frac{1}{\eta_1}\mathbf{I}_{n_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \frac{1}{\eta_2}\mathbf{I}_{n_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \frac{1}{\eta_s}\mathbf{I}_{n_s} \end{bmatrix}$$

DF are easy because the design is balanced:

$$\begin{aligned} DF &= p + \sum_{k=1}^{q} (cm\eta_0)/(cm\eta_0 + \phi_k) = p + \sum_{j=1}^{s} n_j (cm\eta_0)/(cm\eta_0 + \eta_j) \\ &= p + \sum_{j=1}^{s} n_j cm/(cm + r_j), \text{ for } r_j = \eta_j/\eta_0 \\ &= p + \sum_{j=1}^{s} \rho_j, \text{ where } \rho_j = \text{DF controlled by } \eta_j. \end{aligned}$$

A prior on η completes the specification

Unconditional priors:

- $\eta_j \sim$ Gamma, independently

  Note: $\Gamma(0.001, 0.001)$ has mean 1, var 1000, 95^{th} %ile 3×10^{-20} (!)

- $\rho_j \in [0, n_j]$ gets a prior (DF controlled by η_j)
  - flat prior is proper (= uniform-shrinkage prior)
  - $\rho_j/n_j \sim$ Beta(a, b), mean $a/(a+b)$, var $ab/(a+b)^2(a+b+1)$
  - Cui et al (2010): These priors are invariant to re-parameterizations.

- Two-point prior:
  - $P(\rho_j = \xi) = 0.5$, for small $\xi > 0$,
  - $P(\rho_j = n_j - \xi) = 0.5$

Prior on η, conditional on DF

Condition on $\rho_j = K$: This $\Rightarrow r_j = \eta_j/\eta_0 = cm(n_j - K)/K$ is fixed.

Condition on $\sum_{j \in S} \rho_j = K$ or $\leq K$, for S a subset of $\{1, \ldots, s\}$:

This does not fix any of the $\rho_j, j \in S$.

It fixes the total complexity of $\{\theta_k | j(k) \in S\}$ at K or $\leq K$ DF.

The groups of θ_k, for $j \in S$, compete for the K DF.

One way to specify such a prior:

- Specify unconditional priors on $\{\rho_j\}$ or $\{r_j\}$ or $\{\eta_j\}$,
- Then impose the condition.

The prior on DF *does* matter (HCSC 2007 *Technometrics*)

The two-point prior performs poorly if error variation is substantial:
SANOVA + two-point prior = guess whether to include each effect, and the machinery guesses wrong frequently.

These other priors performed similarly as unconditional priors:

- gamma on η_j
- flat on ρ_j
- Beta(0.5, 0.5) on ρ_j.

Analysis of the polishability data

Next slide: Main effect design matrix columns (NOT scaled).

I show three analyses, all using these priors:

- M×P and M×F interactions: each θ_k has its own η_j.
- P×F interaction: smooth its 21 contrasts using a single η_j.
- A flat prior on η_0 (error precision).
- A flat prior on all ρ_j, including the M×P×F interaction.

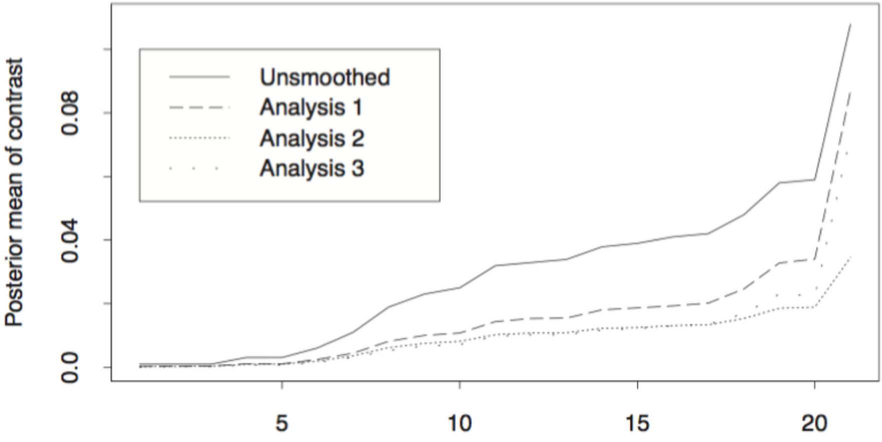
The analyses differ in the prior on the M×P×F interaction:

- Analysis 1: Each of M×P×F's 21 contrasts has its own η_j; they're smoothed separately.
- Analysis 2: Smooth all 21 contrasts using one η_j; $E(DF|\mathbf{y}) = 6.75$.
- Analysis 3: Each of M×P×F's 21 contrasts has its own η_j, but condition the prior on M×P×F's total DF $= 6.75$.

gap (μm)	Material	Polishing			Finishing						
5.02	1	3	0	0	7	0	0	0	0	0	0
8.84	1	3	0	0	−1	6	0	0	0	0	0
3.61	1	3	0	0	−1	−1	5	0	0	0	0
10.55	1	3	0	0	−1	−1	−1	4	0	0	0
3.90	1	3	0	0	−1	−1	−1	−1	3	0	0
5.64	1	3	0	0	−1	−1	−1	−1	−1	2	0
98.95	1	3	0	0	−1	−1	−1	−1	−1	−1	1
10.75	1	3	0	0	−1	−1	−1	−1	−1	−1	−1
2.91	1	−1	2	0	7	0	0	0	0	0	0
3.00	1	−1	2	0	−1	6	0	0	0	0	0
5.94	1	−1	2	0	−1	−1	5	0	0	0	0
8.64	1	−1	2	0	−1	−1	−1	4	0	0	0
16.33	1	−1	2	0	−1	−1	−1	−1	3	0	0
7.44	1	−1	2	0	−1	−1	−1	−1	−1	2	0
11.26	1	−1	2	0	−1	−1	−1	−1	−1	−1	1
16.35	1	−1	2	0	−1	−1	−1	−1	−1	−1	−1
4.75	1	−1	−1	1	7	0	0	0	0	0	0
3.93	1	−1	−1	1	−1	6	0	0	0	0	0
4.90	1	−1	−1	1	−1	−1	5	0	0	0	0
13.44	1	−1	−1	1	−1	−1	−1	4	0	0	0
2.82	1	−1	−1	1	−1	−1	−1	−1	3	0	0
6.44	1	−1	−1	1	−1	−1	−1	−1	−1	2	0
20.88	1	−1	−1	1	−1	−1	−1	−1	−1	−1	1
9.30	1	−1	−1	1	−1	−1	−1	−1	−1	−1	−1
178.22	1	−1	−1	−1	7	0	0	0	0	0	0
1.95	1	−1	−1	−1	−1	6	0	0	0	0	0
3.70	1	−1	−1	−1	−1	−1	5	0	0	0	0
18.11	1	−1	−1	−1	−1	−1	−1	4	0	0	0
16.40	1	−1	−1	−1	−1	−1	−1	−1	3	0	0
9.61	1	−1	−1	−1	−1	−1	−1	−1	−1	2	0
36.52	1	−1	−1	−1	−1	−1	−1	−1	−1	−1	1
14.88	1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
18.68	−1	3	0	0	7	0	0	0	0	0	0
49.02	−1	3	0	0	−1	6	0	0	0	0	0
4.55	−1	3	0	0	−1	−1	5	0	0	0	0

M×P×F interaction: contrasts, unshrunk and shrunken

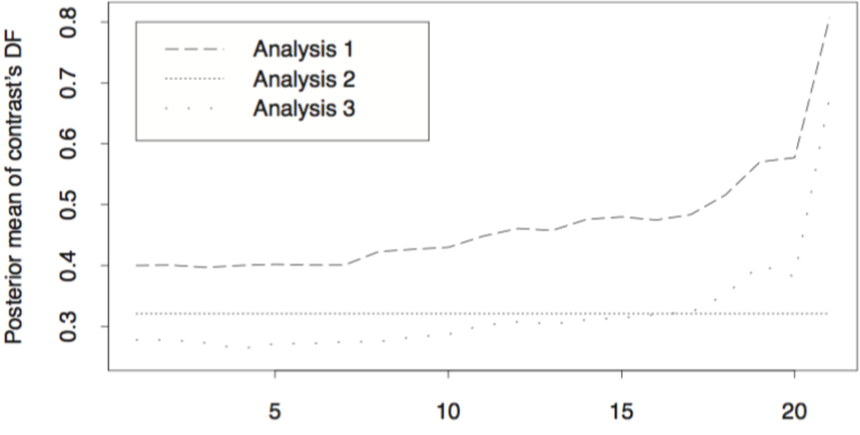
The plot shows absolute values of posterior mean for the 21 contrasts.



(b) Contrasts sorted by absolute value of unsmoothed estimate

M×P×F interaction: $E(\rho_j|\mathbf{y})$

The plot shows posterior mean of DF in each of the 21 contrasts.



(a) Contrasts sorted by absolute value of unsmoothed estimate

Observations about the two preceding plots

Analysis 1:

The outlier pulls up the right-most contrast; it has $E(DF|\mathbf{y}) = 0.8$ DF.

The left-most contrasts have $E(DF|\mathbf{y}) \approx 0.4$ *versus* prior mean 0.5.

Analysis 2:

All 21 contrasts have $E(DF|\mathbf{y}) = 0.32$.

All are smoothed more than in Analysis 1; $E(DF|\mathbf{y}) = 6.75$ total.

The most striking effect is on the largest contrast.

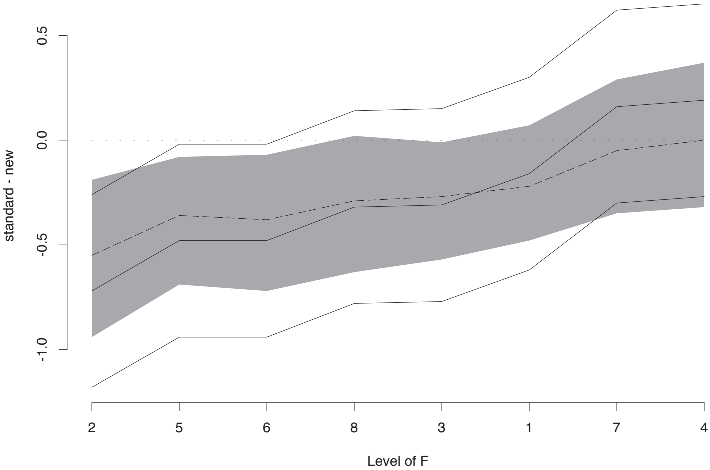
Analysis 3:

Conditioning on 6.75 DF "unmasks" the outlier.

Subgroup analysis: New vs. old for the 8 finishing methods

Solid lines: Unsmoothed estimates and 95% CI for standard minus new.

Dashed line & gray: Analysis 2 estimates and 95% posterior intervals.



Observations about the preceding plot

Analyses 1, 2, and 3 give nearly identical posterior means for the M$\times$F interaction.

I've shown Analysis 2, which gave the widest posterior intervals.

The 7 contrasts in the M$\times$F interaction are smoothed by different $\eta_j \Rightarrow$ shrinkage differs for the levels of F.