

1 Stat 511 Fix-up of the Type II Sum of Squares Stuff in 2-Way Factorials (Plus a Bit)

Thanks go to Norma and Stuart Gardner for pointing out that I need to straighten this out (and telling me what the SAS documentation claims). I goofed up the Type II Sum of Squares stuff when I presented it in class. Here is the correct version.

“Source”	Type I SS (in the order A,B,A × B)	Type II SS	Type III SS
A	$R(\alpha^*{}^{\prime}s \mu^*)$	$R(\alpha^*{}^{\prime}s \mu^*,\beta^*{}^{\prime}s)$	SS_{H_0} for $H_0:\alpha_i = 0 \forall i$ in cell means model
B	$R(\beta^*{}^{\prime}s \mu^*,\alpha^*{}^{\prime}s)$	$R(\beta^*{}^{\prime}s \mu^*,\alpha^*{}^{\prime}s)$	SS_{H_0} for $H_0:\beta_j = 0 \forall j$ in cell means model
A×B	$R(\alpha\beta^*{}^{\prime}s \mu^*,\alpha^*{}^{\prime}s,\beta^*{}^{\prime}s)$	$R(\alpha\beta^*{}^{\prime}s \mu^*,\alpha^*{}^{\prime}s,\beta^*{}^{\prime}s)$	SS_{H_0} for $H_0:\alpha\beta_{ij} = 0 \forall i, j$ in cell means model

The “A×B” sums of squares of all three types are the same. *When the data are balanced*, the “A” and “B” sums of squares of all three types are also the same. But when the sample sizes are not the same, the “A” and “B” sums of squares of different “types” are generally not the same, only the Type III sums of squares are appropriate for testing hypotheses $H_0:\alpha_i = 0 \forall i$ or $H_0:\beta_j = 0 \forall j$ in the cell means model. Exactly what hypotheses in the cell means model could be tested using the Type I or Type II sums of squares is both hard to figure out and actually quite bizarre when one does figure it out.

Notice also that perhaps contrary to one’s naive hopes, a hypothesis like $H_0:\alpha_i^* = 0 \forall i$ in a restricted effects model is not necessarily the same hypothesis as $H_0:\alpha_i = 0 \forall i$ in the cell means model. The latter is a hypothesis that row average means are the same for all rows. The former, that could be tested with a numerator sum of squares

$$R(\alpha^*{}^{\prime}s|\mu^*,\beta^*{}^{\prime}s,\alpha\beta^*{}^{\prime}s)$$

(which is what I erroneously told you in class was the Type II sum of squares for Factor A), is **not** necessarily any such hypothesis.

To see this consider a simple 2×3 effects model under the SAS baseline restriction. For this set-up, the 6 cell means are as below

		B		
		1	2	3
A	1	$\mu^* + \alpha_1^* + \beta_1^* + \alpha\beta_{11}^*$	$\mu^* + \alpha_1^* + \beta_2^* + \alpha\beta_{12}^*$	$\mu^* + \alpha_1^*$
	2	$\mu^* + \beta_1^*$	$\mu^* + \beta_2^*$	μ^*

Here the hypothesis that the two row average means are the same is the hypothesis

$$H_0 : \alpha_1^* + \frac{1}{3}(\alpha\beta_{11}^* + \alpha\beta_{12}^*) = 0$$

The hypothesis $H_0: \alpha_1^* = 0$ in this parameterization is the hypothesis that in fact the cell means are as in the table below. That is, the hypothesis is the hypothesis that $\mu_{13} = \mu_{23}$!!!!

		B		
		1	2	3
A	1	$\mu^* + \beta_1^* + \alpha\beta_{11}^*$	$\mu^* + \beta_2^* + \alpha\beta_{12}^*$	μ^*
	2	$\mu^* + \beta_1^*$	$\mu^* + \beta_2^*$	μ^*