Lecture 37 – Making Sense of Degrees of Freedom in ANOVA

Degrees of freedom in analysis of variance can seem arbitrary and confusing. I'm going to attempt to demystify it for the case of a two-way ANOVA with interactions on the theory that if we can understand degrees of freedom for a relatively complex model, simpler models ought to be a piece of cake. The strategy I'll follow centers on finding the degrees of freedom for estimating the variance of the error two different ways: what I'll call the bottom-up and top-down approaches. The goal is to explain the degrees of freedom for sums of squares that appear in the table below.

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
MAIN EFFECTS				
A: (Factor A)	*SSA*	a - 1	*SSA*/(a – 1)	*MSA*/*MSE*
B: (Factor B)	*SSB*	b - 1	*SSB*/(b – 1)	*MSB*/*MSE*
INTERACTIONS				
AB	*SSAB*	(a – 1)(b – 1)	*SSAB*/(a – 1)(b – 1)	*MSAB*/*MSE*
RESIDUAL (Error)	*SSE*	$ab(n_s-1)$	*SSE*/$ab(n_s-1)$	
TOTAL	*SST*	n - 1		

Estimating Parameters in Two-Way ANOVA with Interactions

In two-way ANOVA with interactions, the goal is to estimate the treatment means. (The model assumes variance is constant for all treatment means, but at least some of the means themselves differ.) Throughout this discussion, and in the table above, we are assuming a complete balanced design (all treatments are sampled and all samples are of the same size n_s). Factor A has a levels, and Factor B has b levels.

In the bottom-up approach, we consider how much information is available to estimate the error variance. We start with the n observations, where $n = abn_s$. Within the ij^{th} treatment, there are $n_s - 1$ independent observations available for estimating variance *after* estimating the treatment mean μ_{ij} by $\bar{Y}_{ij} = \frac{1}{n_s}\sum_{k=1}^{n_s} Y_{ijk}$.

Viewed another way, the degrees of freedom for estimating σ^2 by $S_{ij}^2 = \frac{1}{n_s-1}\sum_{k=1}^{n_s}\left(Y_{ijk} - \bar{Y}_{ij}\right)^2$ is $n_s - 1$.

There are ab treatments, so the total degrees of freedom available for estimating the error variance σ^2 from the sample is $ab(n_s-1) = n - ab$.

In the top-down approach, we begin by estimating the overall mean (the mean across all treatments) μ, then estimate the factor level means $\mu_{i\cdot}$ and $\mu_{\cdot j}$, and finally estimate the treatment means μ_{ij}, subtracting the information (degrees of freedom) "used up" estimating the means as we go along. Whatever remains is available for estimating the error variance σ^2. The tables below should help visualize the process.

The result is that $df_{SSE} = n - 1 - (a-1) - (b-1) - (a-1)(b-1) = n - ab$, as in the bottom-up approach.

Factor A : Factor B	1	2	$\cdots$	*b*-1	*b*	Factor A Means
1	$\bar{Y}_{11}$	$\bar{Y}_{12}$		$\bar{Y}_{1(b-1)}$		$\bar{Y}_{1\cdot}$
2	$\bar{Y}_{21}$	$\bar{Y}_{22}$		$\vdots$		$\bar{Y}_{2\cdot}$
$\vdots$				$\vdots$		$\bar{Y}_{i\cdot}$
a-1	$\bar{Y}_{(a-1)1}$	$\cdots$	$\cdots$	$\bar{Y}_{(a-1)(b-1)}$		$\bar{Y}_{(a-1)\cdot}$
a						
Factor B Means	$\bar{Y}_{\cdot 1}$	$\bar{Y}_{\cdot 2}$	$\bar{Y}_{\cdot j}$	$\bar{Y}_{\cdot(b-1)}$		$\bar{\bar{Y}}$

Table of Independent Estimators in the Top-down Approach

		Grand Mean		Factor A Means		Factor B Means		Treatment Means
Parameter		μ		$\mu_{i\cdot}$		$\mu_{\cdot j}$		μ_{ij}
Estimator		$\bar{\bar{Y}}$		$\bar{Y}_{i\cdot}$		$\bar{Y}_{\cdot j}$		$\bar{Y}_{ij}$
Information	=	*n*-1	-	(*a*-1)	-	(*b*-1)	-	(*a*-1)(*b*-a)
df_{SSE}	=	df_{SST}	-	df_{SSA}	-	df_{SSB}	-	df_{SSAB}

Schematic table showing where the original *n* bits of information in the sample go prior to estimating the error variance in the top-down approach. Recall,

$$df_{SST} = df_{SSE} + df_{SSA} + df_{SSB} + df_{SSAB} \,.$$

Estimating Parameters in Two-Way ANOVA with Without Interactions

If interactions aren't significant, the parameters of interest are the factor level means for factors A and B. Then the top-down approach produces a table like the one below (notice the treatment means are absent).

		Grand Mean		Factor A Means		Factor B Means
Parameter		μ		$\mu_{i.}$		$\mu_{.j}$
Estimator		$\bar{\bar{Y}}$		$\bar{Y}_{i.}$		$\bar{Y}_{.j}$
Information	=	n-1	-	(a-1)	-	(b-1)
df_{SSE}	=	df_{SST}	-	df_{SSA}	-	df_{SSB}

Then, $df_{SSE} = n-1-(a-1)-(b-1) = n-a-b+1$. The corresponding ANOVA table appears below.

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
MAIN EFFECTS				
A: (Factor A)	*SSA*	*a* - 1	*SSA*/(*a* – 1)	*MSA*/*MSE*
B: (Factor B)	*SSB*	*b* - 1	*SSB*/(*b* – 1)	*MSB*/*MSE*
RESIDUAL (Error)	*SSE*	*n* – *a* – *b* + 1	*SSE*/(*n* – *a* – *b* + 1)	
TOTAL	*SST*	*n* - 1		