2F Unbalanced Completely Randomized ANOVA

Engineering Statistics II

Josh Engwer

TTU

2019

Unbalanced 2F ANOVA (2F ucrANOVA)

Definition

(Unbalanced Experiment)

An experiment with unequal group sizes is an **unbalanced experiment**.

SYNONYMS: Unbalanced/Non-orthogonal data/design/model/ANOVA

Unbalanced 1F ANOVA (1F ucrANOVA) was covered in §10.3 of this course.

FACTOR B: $\rightarrow$ **FACTOR A:** $\downarrow$	Level 1 $(x_{\bullet 1})$	Level 2 $(x_{\bullet 2})$
Level 1 $(x_{1\bullet})$	$x_{111},\ x_{112}$ $x_{113},\ x_{114}$	$x_{121},\ x_{122}$
Level 2 $(x_{2\bullet})$	$x_{211},\ x_{212}$	$x_{221},\ x_{222},\ x_{223}$
Level 3 $(x_{3\bullet})$	x_{311}	$x_{321},\ x_{322}$

Prototype 3×2 (Factor A $\times$ Factor B) unbalanced experiment

$(x_{ijk} \equiv k^{th}$ measurement at i^{th} level of factor A $\&$ j^{th} level of factor B$)$

Unbalanced 2-Factor ANOVA

Definition

(Unbalanced Experiment)

An experiment with unequal group sizes is an **unbalanced experiment**.

SYNONYMS: Unbalanced/Non-orthogonal data/design/model/ANOVA

Unbalanced 1-factor experiments were handled in §10.3 earlier in the course.

Unfortunately, unbalanced 2-factor experiments are far trickier to analyze:

1. Sums of squares & F stat values are no longer (nearly) independent‡.
2. Partitioning the sums of squares may no longer work†‡.

Balanced 2F Data	$SS_A,\ SS_B,\ SS_{AB},\ SS_{err}$ are independent $f_A,\ f_B,\ f_{AB}$ are nearly independent (for large K) Always, $SS_A + SS_B + SS_{AB} + SS_{err} = SS_{total}$
Unbalanced 2F Data	$SS_A,\ SS_B,\ SS_{AB},\ SS_{err}$ are dependent $f_A,\ f_B,\ f_{AB}$ are dependent Often, $SS_A + SS_B + SS_{AB} + SS_{err} \neq SS_{total}$

Unbalanced 2-Factor ANOVA

So, how does one perform 2-factor ANOVA with unbalanced data?

The immediate solution is frowned upon by Lomax & Hahs-Vaughn[†]:

> (pg 103): "A rather silly approach, and one that we do not condone, is to delete enough data until you have an equal (group size) model."

i.e. Removing measurements to yield a balanced model is a big no-no!

Unbalanced 2-Factor ANOVA

Three viable solutions exist due to Frank Yates in 1934*:

NAME IN SAS/SPSS:	OTHER COMMON NAME:
Type I Sum of Squares	Yates' Sequential Method
Type II Sum of Squares	Yates' Fitting Constants Method
Type III Sum of Squares	Yates' Weighted Square Means Method

*F. Yates, "The Analysis of Multiple Classifications with Unequal Numbers in the Different Classes", *J. American Statistical Association*, **29** (1934), 51-66.

Many advanced stats books recommend using Type III sums of squares[†‡].

However, Øyvind Langsrud disagrees and recommends Type II:

Ø. Langsrud, "ANOVA for Unbalanced Data: Use Type II instead of Type III Sums of Squares", *Statistics and Computing*, **13** (2003), 163-167.

Anyway, unbalanced 2-factor ANOVA will never be considered in this course.

Papers on History of 2F ucrANOVA

First paper on 2F ucrANOVA:

F. Yates, “The Analysis of Multiple Classifications with Unequal Number in the Different Classes”, *J. American Statistical Association*, **29** (1934), 51-66.

Consult the following papers for more about 2F ucrANOVA:

Y. Fujikoshi, “Two-way ANOVA Models with Unbalanced Data”, *Discrete Mathematics*, **116** (1993), 315-334.

D.G. Herr, “On the History of ANOVA of Unbalanced, Factorial Designs: The First 30 Years”, *The American Statistician*, **40** (1986), 265-270.

†R.G. Lomax, D.L. Hahs-Vaughn, *Statistical Concepts: A* 2^{nd} *Course*, 4^{th} Ed, Routledge, 2012.

‡J.P. Stevens, *Intermediate Statistics: A Modern Approach*, 3^{rd} Ed, Taylor & Francis, 2007.

Fin

Fin.