ORIGIN ≡ 0

W. Stein

Two-Way ANOVA - Equal Sample Sizes

The ANOVA approach analyzes means from multiple populations with membership in each sample determined by discrete values of a classification variable. The Two-Way (and higher) ANOVA stategy extends the system of classifications to two (or more) variables. Here we look at analysis of fully randomized ballanced designs in which numbers of observations in each class (or block) of data are all the same.

Data Structure:

Data are structured as an R X C Contingency Table with cells representing simultaneous classification by two variables. Numeric values Yij for n objects are placed in each cell

Let index i,j indicate the ith row (treatment classes of Variable R) and jth column (treatment classes of Variable C)

Treatment Classes of Variable A:	Two-Way ANOVA				
	Treatment Classes of Variable B:				
	#1	#2	#3	...	#j
#1	n	n	n	...	n
#2	n	n	n	...	n
#3	n	n	n	...	n
...	...	...	...	...	...
#i	n	n	n	...	n
Each cell consists of n replicates with means $Ybar_{ij}$					

Also let: $A_{i.}$ **= mean over all columns for row i.**
$B_{j.}$ **= mean over all rows for column j.**
GM = overall mean.

Model:

$Y_{i,j} = \mu_{ij} + \varepsilon_{ijk}$ < cell means model

$Y_{i,j} = \mu.. + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$ < factor effects model

where:
μ **is a constant = grand mean of all objects.**
α_i **is effect coefficient for classes i in Variable A.**
β_j **is effect coefficient for classes j in Variable B.**
$\alpha\beta_{ij}$ **is interaction coefficient for classes i,j between Variables A and B.**
ε_{ijk} **is the error term specific to each object i,j,k**

Restrictions:

$\sum_i \alpha_i := 0 \quad \sum_j \beta_j := 0 \quad \sum_i \alpha\beta_{ij} := 0$ **for all i & j**

Assumptions:

– ε_{ijk} **are a random sample ~ N(0,** σ^2**)**
– **variance is homogeneous across cells**

Example:

K := READPRN("c:/2008LinearModelsData/CastleBakeryR.txt")

Variable Assignment:

	B_1	B_2
A_1	$Y_{11} := \begin{pmatrix} 47 \\ 43 \end{pmatrix}$	$Y_{12} := \begin{pmatrix} 46 \\ 40 \end{pmatrix}$
A_2	$Y_{21} := \begin{pmatrix} 62 \\ 68 \end{pmatrix}$	$Y_{22} := \begin{pmatrix} 67 \\ 71 \end{pmatrix}$
A_3	$Y_{31} := \begin{pmatrix} 41 \\ 39 \end{pmatrix}$	$Y_{32} := \begin{pmatrix} 42 \\ 46 \end{pmatrix}$

$Y := K^{\langle 1 \rangle}$

n := 2

a := 3

b := 2

$$Y = \begin{pmatrix} 47 \\ 43 \\ 46 \\ 40 \\ 62 \\ 68 \\ 67 \\ 71 \\ 41 \\ 39 \\ 42 \\ 46 \end{pmatrix} \qquad K = \begin{pmatrix} 1 & 47 & 1 & 1 & 1 \\ 2 & 43 & 1 & 1 & 2 \\ 3 & 46 & 1 & 2 & 1 \\ 4 & 40 & 1 & 2 & 2 \\ 5 & 62 & 2 & 1 & 1 \\ 6 & 68 & 2 & 1 & 2 \\ 7 & 67 & 2 & 2 & 1 \\ 8 & 71 & 2 & 2 & 2 \\ 9 & 41 & 3 & 1 & 1 \\ 10 & 39 & 3 & 1 & 2 \\ 11 & 42 & 3 & 2 & 1 \\ 12 & 46 & 3 & 2 & 2 \end{pmatrix}$$

Means:

$GM := mean(Y)$ $GM = 51$ **< grand mean**

$$A := \begin{pmatrix} mean(Y_{11}, Y_{12}) \\ mean(Y_{21}, Y_{22}) \\ mean(Y_{31}, Y_{32}) \end{pmatrix} \qquad A = \begin{pmatrix} 44 \\ 67 \\ 42 \end{pmatrix}$$

< factor A means

$$B := \begin{pmatrix} mean(Y_{11}, Y_{21}, Y_{31}) \\ mean(Y_{12}, Y_{22}, Y_{32}) \end{pmatrix} \qquad B = \begin{pmatrix} 50 \\ 52 \end{pmatrix}$$

< factor B means

$$AB := \begin{pmatrix} mean(Y_{11}) & mean(Y_{12}) \\ mean(Y_{21}) & mean(Y_{22}) \\ mean(Y_{31}) & mean(Y_{32}) \end{pmatrix} \qquad AB = \begin{pmatrix} 45 & 43 \\ 65 & 69 \\ 40 & 44 \end{pmatrix}$$

^ block means

$$W := \begin{pmatrix} mean(Y_{11}) \\ mean(Y_{11}) \\ mean(Y_{12}) \\ mean(Y_{12}) \\ mean(Y_{21}) \\ mean(Y_{21}) \\ mean(Y_{22}) \\ mean(Y_{22}) \\ mean(Y_{31}) \\ mean(Y_{31}) \\ mean(Y_{32}) \\ mean(Y_{32}) \end{pmatrix} \qquad W = \begin{pmatrix} 45 \\ 45 \\ 43 \\ 43 \\ 65 \\ 65 \\ 69 \\ 69 \\ 40 \\ 40 \\ 44 \\ 44 \end{pmatrix}$$

^ block means in vector form

Sums of Squares:

SSA for factor A:

$i := 0 .. length(A) - 1$

$$SSA := n \cdot b \cdot \sum_i (A_i - GM)^2 \qquad SSA = 1544$$

SSB for factor B:

$j := 0 .. length(B) - 1$

$$SSB := n \cdot a \cdot \sum_j (B_j - GM)^2 \qquad SSB = 12$$

SSAB for Interaction AB:

$$SSAB := n \cdot \sum_i \sum_j (AB_{i,j} - A_i - B_j + GM)^2 \qquad SSAB = 24$$

SSE for Error:

$k := 0 .. length(W) - 1$

$$SSE := \sum_k (Y_k - W_k)^2 \qquad SSE = 62$$

SSTR for Treatment:

$$SSTR := n \cdot \left[\sum_i \sum_j (AB_{i,j} - GM)^2\right] \qquad SSTR = 1580 \qquad SSA + SSB + SSAB = 1580$$

SSTO for Total:

$$SSTO := \sum_k (Y_k - GM)^2 \qquad SSTO = 1642 \qquad SSTR + SSE = 1642$$

ANOVA Table for Two-way design:

Sum of Squares:	Degrees of Freedom:		Mean Squares:	
$SSA = 1544$	$df_A := a - 1$	$df_A = 2$	$MSA := \frac{SSA}{df_A}$	$MSA = 772$
$SSB = 12$	$df_B := (b - 1)$	$df_B = 1$	$MSB := \frac{SSB}{df_B}$	$MSB = 12$
$SSAB = 24$	$df_{AB} := (a - 1) \cdot (b - 1)$	$df_{AB} = 2$	$MSAB := \frac{SSAB}{df_{AB}}$	$MSAB = 12$
$SSE = 62$	$df_E := a \cdot b \cdot (n - 1)$	$df_E = 6$	$MSE := \frac{SSE}{df_E}$	$MSE = 10.3333$
$SSTO = 1642$	$df_T := n \cdot a \cdot b - 1$	$df_T = 11$	**^ values verified KNNL p. 842**	

F-Tests in Two-Way ANOVA with Fixed Effects Model:

F-Test for H_0: All $\alpha\beta_{ij} = 0$

Hypotheses:

H_0: $\alpha\beta_{ij} = 0$ *for all* ij **< All interactions between the two variables is 0**

H_1: *At least one* γ_{ij} <> 0

Test Statistic:

$$F_{AB} := \frac{MSAB}{MSE} \qquad F_{AB} = 1.1613$$

Critical Value of the Test:

$\alpha := 0.05$ **< Probability of Type I error must be explicitly set**

$$CV := qF(1 - \alpha, df_{AB}, df_E) \qquad CV = 5.1433$$

Decision Rule:

IF F_{AB} > C, THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability Value:

$$P_{AB} := 1 - pF(F_{AB}, df_{AB}, df_E) \qquad P_{AB} = 0.3747$$

Note: If the results of this test suggest no significant interaction, then tests for main effects (below) have meaning.

F-Test for H_0: All $\alpha_i = 0$

Hypotheses:

H_0: $\alpha_i = 0$ *for all* i **< All treatment class deviations in factor A from the grand mean are 0**

H_1: *At least one* α_i <> 0

Test Statistic:

$$F_A := \frac{MSA}{MSE} \qquad F_A = 74.7097$$

Critical Value of the Test:

$\alpha := 0.05$ **< Probability of Type I error must be explicitly set**

$$CV := qF(1 - \alpha, df_A, df_E) \qquad CV = 5.1433$$

Decision Rule:

IF F_1 > C, THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability Value:

$$P_A := 1 - pF(F_A, df_A, df_E) \qquad P_A = 5.7536 \times 10^{-5}$$

F-Test for H_0: All $\beta_j = 0$

Hypotheses:

H_0: $\beta_j = 0$ *for all* j **< All treatment class deviations in factor B from the grand mean are 0**

H_1: *At least one* β_j <> 0

Test Statistic:

$$F_B := \frac{MSB}{MSE} \qquad F_B = 1.1613$$

Critical Value of the Test:

$\alpha := 0.05$ **< Probability of Type I error must be explicitly set**

$$CV := qF(1 - \alpha, df_B, df_E) \qquad CV = 5.9874$$

Decision Rule:

IF F_2 > C, THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability Value:

$$P_B := 1 - pF(F_B, df_B, df_E) \qquad P_B = 0.3226$$

Prototype in R:

```
#BALANCED 2-WAY ANOVA
require(car)
#READ STRUCTURED DATA TABLE WITH NUMERIC CODED FACTOR
K=read.table("c:/2008LinearModelsData/CastleBakeryR.txt")
K
attach(K)
Y=sales
A=factor(height)
B=factor(width)
detach(K)
contrasts(A)=contr.sum
contrasts(B)=contr.sum
FM=lm(Y~A+B+A:B)
summary(FM)
anova(FM)
Anova(FM, type="3")
```

```
> K
   sales height width replicate
1     47      1     1         1
2     43      1     1         2
3     46      1     2         1
4     40      1     2         2
5     62      2     1         1
6     68      2     1         2
7     67      2     2         1
8     71      2     2         2
9     41      3     1         1
10    39      3     1         2
11    42      3     2         1
12    46      3     2         2
```

```
> summary(FM)
Call:
lm(formula = Y ~ A + B + A:B)
Residuals:
       Min         1Q     Median         3Q        Max
-3.000e+00 -2.000e+00  1.110e-16  2.000e+00  3.000e+00
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   51.000      0.928  54.959 2.44e-09 ***
A1            -7.000      1.312  -5.334  0.00177 **
A2            16.000      1.312  12.192 1.85e-05 ***
B1            -1.000      0.928  -1.078  0.32261
A1:B1          2.000      1.312   1.524  0.17835
A2:B1         -1.000      1.312  -0.762  0.47494
```

```
> anova(FM)
Analysis of Variance Table
Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)
A          2 1544.00  772.00 74.7097 5.754e-05 ***
B          1   12.00   12.00  1.1613    0.3226
A:B        2   24.00   12.00  1.1613    0.3747
Residuals  6   62.00   10.33
```

anova() & Anova{car} > give the same result

```
> Anova(FM, type="3")
Anova Table (Type III tests)
Response: Y
            Sum Sq Df   F value    Pr(>F)
(Intercept)  31212  1 3020.5161 2.437e-09 ***
A             1544  2   74.7097 5.754e-05 ***
B               12  1    1.1613    0.3226
A:B             24  2    1.1613    0.3747
Residuals       62  6
```

$$\begin{pmatrix} SSA & df_A & MSA & F_A & P_A \\ SSB & df_B & MSB & F_B & P_B \\ SSAB & df_{AB} & MSAB & F_{AB} & P_{AB} \end{pmatrix} = \begin{pmatrix} 1544 & 2 & 772 & 74.7097 & 5.7536 \times 10^{-5} \\ 12 & 1 & 12 & 1.1613 & 0.3226 \\ 24 & 2 & 12 & 1.1613 & 0.3747 \end{pmatrix}$$

< ^ same results