ORIGIN ≡ 1

Two-Way ANOVA for Fixed Effects with 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 Fixed Effects (Model I ANOVA) analysis of a *balanced* design in which numbers of observations in each class (or block) of data are all the same. Unbalanced designs, in which the numbers of observations in each group are not the same, require a different approach because sums of squares in the ANOVA are not orthogonal - i.e., do not add as shown below. Instead, use a General Linear Model (GLM) approach.

Data Structure:

Data are structured as an R X C Table with cells representing simultaneous classification by two variables. Numeric values Y_{ij} 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:

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

Assumptions:

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

ZAR := READPRN("c:/DATA/Biostatistics/ZarEX12.1R.txt") **< Zar Example 12.1**

$Y := ZAR^{\langle 2\rangle}$

Variable Assignment:

	calcium	treatment	sex
1	16.3	nohorm	female
2	20.4	nohorm	female
3	12.4	nohorm	female
4	15.8	nohorm	female
5	9.5	nohorm	female
6	15.3	nohorm	male
7	17.4	nohorm	male
8	10.9	nohorm	male
9	10.3	nohorm	male
10	6.7	nohorm	male
11	38.1	horm	female
12	26.2	horm	female
13	32.3	horm	female
14	35.8	horm	female
15	30.2	horm	female
16	34	horm	male
17	22.8	horm	male
18	27.8	horm	male
19	25	horm	male
20	29.3	horm	male

B_1 **B_2**

A_1 $Y_{11} := \begin{pmatrix} 16.3 \\ 20.4 \\ 12.4 \\ 15.8 \\ 9.5 \end{pmatrix}$ $Y_{12} := \begin{pmatrix} 15.3 \\ 17.4 \\ 10.9 \\ 10.3 \\ 6.7 \end{pmatrix}$

A_2 $Y_{21} := \begin{pmatrix} 38.1 \\ 26.2 \\ 32.3 \\ 35.8 \\ 30.2 \end{pmatrix}$ $Y_{22} := \begin{pmatrix} 34 \\ 22.8 \\ 27.8 \\ 25 \\ 29.3 \end{pmatrix}$

$Y := ZAR^{\langle 2 \rangle}$

$n := 5$

$a := 2$

$b := 2$

$N := 20$

Means:

$GM := \text{mean}(Y)$ $GM = 21.825$ **< grand mean**

$A := \begin{pmatrix} \text{mean}(Y_{11}, Y_{12}) \\ \text{mean}(Y_{21}, Y_{22}) \end{pmatrix}$ $A = \begin{pmatrix} 13.5 \\ 30.15 \end{pmatrix}$ **< factor A means**

$B := \begin{pmatrix} \text{mean}(Y_{11}, Y_{21}) \\ \text{mean}(Y_{12}, Y_{22}) \end{pmatrix}$ $B = \begin{pmatrix} 23.7 \\ 19.95 \end{pmatrix}$ **< factor B means**

$AB := \begin{pmatrix} \text{mean}(Y_{11}) & \text{mean}(Y_{12}) \\ \text{mean}(Y_{21}) & \text{mean}(Y_{22}) \end{pmatrix}$ $AB = \begin{pmatrix} 14.88 & 12.12 \\ 32.52 & 27.78 \end{pmatrix}$

^ cell means

$Y = \begin{pmatrix} 16.3 \\ 20.4 \\ 12.4 \\ 15.8 \\ 9.5 \\ 15.3 \\ 17.4 \\ 10.9 \\ 10.3 \\ 6.7 \\ 38.1 \\ 26.2 \\ 32.3 \\ 35.8 \\ 30.2 \\ 34 \\ 22.8 \\ 27.8 \\ 25 \\ 29.3 \end{pmatrix}$ $W := \begin{pmatrix} 14.88 \\ 14.88 \\ 14.88 \\ 14.88 \\ 14.88 \\ 12.12 \\ 12.12 \\ 12.12 \\ 12.12 \\ 12.12 \\ 32.52 \\ 32.52 \\ 32.52 \\ 32.52 \\ 32.52 \\ 27.78 \\ 27.78 \\ 27.78 \\ 27.78 \\ 27.78 \end{pmatrix}$

^ each Y_{ij} stacked **^ within cell means for each Y_{ij}**

Sums of Squares:

SSA for factor A:

$i := 1..2$

$SSA := n \cdot b \cdot \sum_i (A_i - GM)^2$ $SSA = 1386.1125$

SSB for factor B:

$j := 1..2$

$SSB := n \cdot a \cdot \sum_j (B_j - GM)^2$ $SSB = 70.3125$

SSAB for Interaction AB:

$SSAB := n \cdot \sum_i \sum_j (AB_{i,j} - A_i - B_j + GM)^2$ $SSAB = 4.9005$

SSE for Error:

$k := 1 .. N$

$SSE := \sum_k (Y_k - W_k)^2$ $SSE = 301.392$

SS Total:

$SSTO := \sum_k (Y_k - GM)^2$ $SSTO = 1762.7175$

ANOVA Table for Two-way design:

Sum of Squares:	Degrees of Freedom:		Mean Squares:	
$SSA = 1386.1125$	$df_A := a - 1$	$df_A = 1$	$MSA := \frac{SSA}{df_A}$	$MSA = 1386.1125$
$SSB = 70.3125$	$df_B := (b - 1)$	$df_B = 1$	$MSB := \frac{SSB}{df_B}$	$MSB = 70.3125$
$SSAB = 4.9005$	$df_{AB} := (a - 1) \cdot (b - 1)$	$df_{AB} = 1$	$MSAB := \frac{SSAB}{df_{AB}}$	$MSAB = 4.9005$
$SSE = 301.392$	$df_E := a \cdot b \cdot (n - 1)$	$df_E = 16$	$MSE := \frac{SSE}{df_E}$	$MSE = 18.837$
$SSTO = 1762.7175$	$df_T := n \cdot a \cdot b - 1$	$df_T = 19$		

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* $\alpha\beta_{ij}$ <> 0

Test Statistic:

$F_{AB} := \frac{MSAB}{MSE}$ $F_{AB} = 0.2602$

Sampling Distribution of the Test Statistic F:

If Assumptions hold and H_0 is true then F ~$F_{(dfAB,dfE)}$

Critical Value of the Test:

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

$C := qF(1 - \alpha, df_{AB}, df_E)$ $C = 4.494$

Decision Rule:

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

Probability Value:

$P_{AB} := 1 - pF(F_{AB}, df_{AB}, df_E)$ $P_{AB} = 0.616979$

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

$F_A := \frac{MSA}{MSE}$ $F_A = 73.5846$

Sampling Distribution of the Test Statistic F:

If Assumptions hold and H_0 is true then F ~$F_{(dfA,dfE)}$

Critical Value of the Test:

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

$C_A := qF(1 - \alpha, df_A, df_E)$ $C_A = 4.494$

Decision Rule:

IF F_1 > C, THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability Value:

$P_A := 1 - pF(F_A, df_A, df_E)$ $P_A = 2.2172 \times 10^{-7}$

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:

$F_B := \frac{MSB}{MSE}$ $F_B = 3.7327$

Sampling Distribution of the Test Statistic F:

If Assumptions hold and H_0 is true then F ~$F_{(dfB,dfE)}$

Critical Value of the Test:

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

$C_B := qF(1 - \alpha, df_B, df_E)$ $C_B = 4.494$

Decision Rule:

IF $F_2 > C$, THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability Value:

$P_B := 1 - pF(F_B, df_B, df_E)$ $P_B = 0.071264$

Prototype in R:

```
#ZAR EXAMPLE 12.1
ZAR=read.table("c:/DATA/Biostatistics/ZarEX12.1R.txt")
ZAR
attach(ZAR)
A=factor(treatment)
B=factor(sex)
Y=calcium
options(digits=9)
```

```
Analysis of Variance Table

Response: Y
          Df    Sum Sq   Mean Sq  F value      Pr(>F)
A          1 1386.1125 1386.1125 73.58457 2.2172e-07 ***
B          1   70.3125   70.3125  3.73268   0.071264 .
A:B        1    4.9005    4.9005  0.26015   0.616979
Residuals 16  301.3920   18.8370
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lm(Y~A+B+A:B)) #FACTORS A,B PLUS A:B INTERACTION
anova(lm(Y~A*B))  # SIMPLER EQUIVALENT FORMULA
```

Confidence Intervals for Factor Means:

Critical Value:

$C := \left| qt\left(\frac{\alpha}{2}, df_E\right) \right|$ $C = 2.1199$

For Factor A:

$CI_A := \left(A - C \cdot \sqrt{\frac{MSE}{b \cdot n}} \quad A + C \cdot \sqrt{\frac{MSE}{b \cdot n}} \right)$ $CI_A = \left[\begin{pmatrix} 10.5905 \\ 27.2405 \end{pmatrix} \begin{pmatrix} 16.4095 \\ 33.0595 \end{pmatrix} \right]$

For Factor B:

$CI_B := \left(B - C \cdot \sqrt{\frac{MSE}{a \cdot n}} \quad B + C \cdot \sqrt{\frac{MSE}{a \cdot n}} \right)$ $CI_B = \left[\begin{pmatrix} 20.7905 \\ 17.0405 \end{pmatrix} \begin{pmatrix} 26.6095 \\ 22.8595 \end{pmatrix} \right]$

^ Note: Zar's calculation of standard error in Example 12.3 appears to be in error in both cases.

From calculations above:

$A = \begin{pmatrix} 13.5 \\ 30.15 \end{pmatrix}$ **< factor A means**

$B = \begin{pmatrix} 23.7 \\ 19.95 \end{pmatrix}$ **< factor B means**

$df_A = 1$ $a = 2$

$df_B = 1$ $b = 2$

$df_{AB} = 1$ $n = 5$

$N = 20$

$df_E = 16$ $MSE = 18.837$

Pooling values across Factor B (sex):

$B_p := \text{mean}(B)$ $B_p = 21.825$

$CI_{Bp} := \left(B_p - C \cdot \sqrt{\frac{MSE}{N}} \quad B_p + C \cdot \sqrt{\frac{MSE}{N}} \right)$ $CI_{Bp} = (19.7677 \quad 23.8823)$

^ this agrees with Zar.