ORIGIN ≡ 0

W. Stein

One-Way Analysis of Variance with Fixed Effects Model

Analysis of Variance (ANOVA) are a broad class of statistical models that fall under the GLM framework. However unlike typical regression where all variables are typically continuous, the independent variables in ANOVA involve membership identifiers for classes. Since more than two classes may be present, this approach allows extension of the t-test strategy to comparisions of multiple populations. Since ANOVA is ubiquitous in many experimental settings in biology, its proficient use is often viewed as evidence of good experimental design. Example below comes from Chapter 16 in Kuter et al. (KNNL) A*pplied Linear Statistical Models* 5th Edition.

Data Structure:

r groups with not necessarily the same numbers of observations and different means.

Let index i,j indicate the ith column (treatment class) and jth row (object or case).

One-Way ANOVA					
	Treatment Classes:				
Objects (Replicates)	#1	#2	#3	…	#r
1					
2					
3					
…					
n	n1	n2	n3		nr
means:	Ybar.1	Ybar.2	Ybar.3	…	Ybar.r

Example:

Kenton Food Exampl KNNL Table 16.1

Cell Means ANOVA Model: **KNNL p.683**

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

< original variable called "Design" is the independent factor with r=4 levels.

$Y := K^{\langle 0 \rangle}$

		K =		Y =
11	1	1		11
17	1	2		17
16	1	3		16
14	1	4		14
15	1	5		15
12	2	1		12
10	2	2		10
15	2	3		15
19	2	4		19
11	2	5		11
23	3	1		23
20	3	2		20
18	3	3		18
17	3	4		17
27	4	1		27
33	4	2		33
22	4	3		22
26	4	4		26
28	4	5		28

Cell Means Model:

$Y_{i,j} = \mu_i + \varepsilon_{i,j}$

Factor Effects Model:

$Y_{i,j} = \mu. + \tau_i + \varepsilon_{i,j}$

Equivalent Models, where:
$\mu.$ is the grand mean of all objects
μ_i is mean of each treatment class i (cell or block)
τ_i is the treatment effect = $\mu. - \mu_i$ for each class i.
$\varepsilon_{i,j}$ is the error term specific to each object i,j
i = 1 to r, j = 1 to n_r for each treatment class i

Assumption:

ε_{ij} are a random sample ~ $N(0,\sigma^2)$

Restriction:

$\Sigma\tau_i = 0$ **< allows estimation of r parameters. Other restrictions are also possible...**

Defining Treatment Classes:

$i := 0..4$ $\quad Y1_i := Y_i$ $\quad Y2_i := Y_{i+5}$ $\quad Y4_i := Y_{i+14}$

$i := 0..3$ $\quad Y3_i := Y_{i+10}$

$$Y_1 = \begin{pmatrix} 11 \\ 17 \\ 16 \\ 14 \\ 15 \end{pmatrix} \qquad Y_2 = \begin{pmatrix} 12 \\ 10 \\ 15 \\ 19 \\ 11 \end{pmatrix} \qquad Y_3 = \begin{pmatrix} 23 \\ 20 \\ 18 \\ 17 \end{pmatrix} \qquad Y_4 = \begin{pmatrix} 27 \\ 33 \\ 22 \\ 26 \\ 28 \end{pmatrix}$$

< Treatment classes are vectors

Number & Means:

$n := \text{length}(Y)$ $\quad n = 19$ **< total number of observations**

$n_1 := \text{length}(Y_1)$ $\quad n_1 = 5$

$n_2 := \text{length}(Y_2)$ $\quad n_2 = 5$ **< cell numbers**

$n_3 := \text{length}(Y_3)$ $\quad n_3 = 4$

$n_4 := \text{length}(Y_4)$ $\quad n_4 = 5$

$r := 4$ **< treatment (block or cell) classes**

$GM := \text{mean}\left(K^{\langle 0 \rangle}\right)$ $\quad GM = 18.6316$ **< grand mean - sample estimate of μ.**

$Y_{bar1} := \text{mean}(Y_1)$ $\quad Y_{bar1} = 14.6$

$Y_{bar2} := \text{mean}(Y_2)$ $\quad Y_{bar2} = 13.4$

$Y_{bar3} := \text{mean}(Y_3)$ $\quad Y_{bar3} = 19.5$ **< Treatment means**

$Y_{bar4} := \text{mean}(Y_4)$ $\quad Y_{bar4} = 27.2$

Sums of Squares:

Treatment:

$$SSTR := n_1 \cdot (Y_{bar1} - GM)^2 + n_2 \cdot (Y_{bar2} - GM)^2 + n_3 \cdot (Y_{bar3} - GM)^2 + n_4 \cdot (Y_{bar4} - GM)^2 \qquad SSTR = 588.2211$$

Error:

$$SSE := \sum_{i=0}^{4} \left(Y1_i - Y_{bar1}\right)^2 + \sum_{i=0}^{4} \left(Y2_i - Y_{bar2}\right)^2 + \sum_{i=0}^{3} \left(Y3_i - Y_{bar3}\right)^2 + \sum_{i=0}^{4} \left(Y4_i - Y_{bar4}\right)^2 \qquad SSE = 158.2$$

Total:

$$SSTO := \sum_{i=0}^{n-1} \left(Y_i - GM\right)^2 \qquad SSTO = 746.4211$$

One-Way ANOVA Table:

Source:	**SS**	**df**	**MS**	
Treatment	$SSTR = 588.2211$	$r - 1 = 3$	$MSTR := \frac{SSTR}{r-1}$	$MSTR = 196.0737$
Error	$SSE = 158.2$	$n - r = 15$	$MSE := \frac{SSE}{n-r}$	$MSE = 10.5467$
TOTAL	$SSTO = 746.4211$			

Overall F-Test:

Hypotheses:

H_0: μ_i **the same** ***for all*** **i**

$\tau_i = 0$ ***for all*** **i** **< Alternate Cell means & Treatment Effects formulations of the same hypothesis**

H_1: ***At least one*** μ_i **different**

At least one $\tau_i <> 0$

Test Statistic:

$$F := \frac{MSTR}{MSE} \qquad F = 18.5911$$

Critical Value of the Test:

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

$$CV := qF[1-\alpha,(r-1),(n-r)] \qquad CV = 3.2874$$

Decision Rule:

IF F > C, THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability Value:

$$P := 1 - pF(F, r-1, n-r) \qquad P = 2.585 \times 10^{-5}$$

Partial t/F Tests:

Hypotheses:

H_0: $\mu_i = \mu_j$ ***for specific*** **i & j**

$\tau_i = \tau_j$ ***for specific*** **i & j** **< Alternate Cell means & Treatment Effects formulations of the same hypothesis**

H_1: $\mu_i <> \mu_j$ ***for specific*** **i & j**

$\tau_i <> \tau_j$ ***for specific*** **i & j**

Test Statistic:

$$t := \begin{bmatrix} \dfrac{Y_{bar1} - Y_{bar2}}{\sqrt{MSE \cdot \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \\ \dfrac{Y_{bar1} - Y_{bar3}}{\sqrt{MSE \cdot \left(\dfrac{1}{n_1} + \dfrac{1}{n_3}\right)}} \\ \dfrac{Y_{bar1} - Y_{bar4}}{\sqrt{MSE \cdot \left(\dfrac{1}{n_1} + \dfrac{1}{n_4}\right)}} \end{bmatrix} \qquad t = \begin{pmatrix} 0.5842 \\ -2.2492 \\ -6.1346 \end{pmatrix}$$

< comparing μ_1 vs μ_2

< comparing μ_1 vs μ_3

< comparing μ_1 vs μ_4

Critical Value of the Test:

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

$C := \left| qt\left(\frac{\alpha}{2}, n - r\right) \right|$ $C = 2.1314$ < **Note degrees of freedom = (n-r)**

Decision Rule:

IF |t| > C, THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability Value:

$i := 0..2$ < **minimum calculated for each test**

$P_i := \min\left[2 \cdot pt\left(t_i, n - r\right), 2 \cdot \left(1 - pt\left(t_i, n - r\right)\right)\right]$ $P = \begin{pmatrix} 0.5677402 \\ 0.0399477 \\ 0.0000191 \end{pmatrix}$

Prototype in R:

```
#READ STRUCTURED DATA TABLE WITH NUMERIC CODED FACTOR
K=read.table("c:/2008LinearModelsData/KentonFoodR.txt")
K
attach(K)
Y=Sales
X=factor(Design) # factor() IN DEFAULT SETTING
FM=lm(Y~X)
summary(FM)
anova(FM)
```

$$\begin{pmatrix} Y_{bar1} \\ Y_{bar2} - Y_{bar1} \\ Y_{bar3} - Y_{bar1} \\ Y_{bar4} - Y_{bar1} \end{pmatrix} = \begin{pmatrix} 14.6 \\ -1.2 \\ 4.9 \\ 12.6 \end{pmatrix}$$

```
> summary(FM)
Call:
lm(formula = Y ~ X)
Residuals:
   Min     1Q Median     3Q    Max 
 -5.20  -1.95  -0.20   1.50   5.80 
Coefficients:
               Estimate  Std. Error  t value   Pr(>|t|)    
(Intercept) 14.60000000  1.45235441 10.05264 4.6632e-08 ***
X2          -1.20000000  2.05393930 -0.58424   0.567740    
X3           4.90000000  2.17853162  2.24922   0.039948 *  
X4          12.60000000  2.05393930  6.13455 1.9101e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
Residual standard error: 3.24756319 on 15 degrees of freedom
Multiple R-squared: 0.788055281,        Adjusted R-squared: 0.745666338 
F-statistic: 18.5910573 on 3 and 15 DF,  p-value: 2.58496098e-05 
> anova(FM)
Analysis of Variance Table
Response: Y
          Df      Sum Sq     Mean Sq  F value     Pr(>F)    
X          3 588.2210526 196.0736842 18.59106 2.5850e-05 ***
Residuals 15 158.2000000  10.5466667                        
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```