People's Democratic Republic of Algeria
University Med Khider of Biskra
Faculty of SNVSTU

Two-way ANOVA

Dr. Ben Gherbal Hanane

Biostatistics — Level: M1 Biology
Email: hanane.bengherbal@univ-biskra.dz

1. Two-way ANOVA with Replication

We study simultaneously two factors A and B:

- p levels for factor A $(A_1, A_2, \ldots, A_p)$, for example: (4 different doses of a certain drug),
- and q levels for factor B $(B_1, B_2, \ldots, B_q)$, for example: (age category: young, adults, elderly).

For each combination (A, B), we have a sample of size n.

	A_1	A_2	....	A_p
B_1	$x_{11,1}, x_{11,2}, \ldots, x_{11,n}$ n observations	$x_{12,1}, x_{12,2}, \ldots, x_{12,n}$ n observations		$x_{1p,1}, x_{1p,2}, \ldots, x_{1p,n}$ n observations
B_2	$x_{21,1}, x_{21,2}, \ldots, x_{21,n}$ n observations	$x_{22,1}, x_{22,2}, \ldots, x_{22,n}$ n observations		$x_{2p,1}, x_{2p,2}, \ldots, x_{2p,n}$ n observations
. . .				
B_q	$x_{q1,1}, x_{q1,2}, \ldots, x_{q1,n}$ n observations	$x_{q2,1}, x_{q2,2}, \ldots, x_{q2,n}$ n observations		$x_{qp,1}, x_{qp,2}, \ldots, x_{qp,n}$ n observations

This is a two-factor crossed ANOVA table with replications.

Assumptions (conditions for application)

- The samples are drawn from Gaussian populations (normal distribution).
  - All populations have the same variance.
  - All samples have the same size.

What we are going to test

- The effect of factor A alone.
- The effect of factor B alone.
- The effect of the interaction between the two factors.

In fact, this test contains 3 sub-tests, therefore we will have 3 null hypotheses and 3 alternative hypotheses.

H_{0A} :	the factor A has no influence on the mean of the populations ($\mu_{A_1} = \mu_{A_2} = \cdots = \mu_{A_p}$)
H_{1A} :	the factor A influences the mean of the populations (there exist at least two different means according to factor A)
H_{0B} :	the factor B has no influence on the mean of the populations ($\mu_{B_1} = \mu_{B_2} = \cdots = \mu_{B_q}$)
H_{1B} :	the factor B influences the mean of the populations (there exist at least two different means according to factor B)
H_{0AB} :	there is no interaction between the two factors.
H_{1AB} :	there is an interaction between the two factors.

Intermediate computations to determine variations

	A_1	A_2	....	A_p	Row means
B_1	$\overline{x}_{11}$	$\overline{x}_{12}$		$\overline{x}_{1p}$	$\overline{x}_{1.} = \frac{1}{p}\sum_{j=1}^{p}\overline{x}_{1j}$
B_2	$\overline{x}_{21}$	$\overline{x}_{22}$		$\overline{x}_{2p}$	$\overline{x}_{2.} = \frac{1}{p}\sum_{j=1}^{p}\overline{x}_{2j}$
. . .					
B_q	$\overline{x}_{q1}$	$\overline{x}_{q2}$		$\overline{x}_{qp}$	$\overline{x}_{q.} = \frac{1}{p}\sum_{j=1}^{p}\overline{x}_{qj}$
Column means	$\overline{x}_{.1} = \frac{1}{q}\sum_{i=1}^{q}\overline{x}_{i1}$	$\overline{x}_{.2} = \frac{1}{q}\sum_{i=1}^{q}\overline{x}_{i2}$		$\overline{x}_{.p} = \frac{1}{q}\sum_{i=1}^{q}\overline{x}_{ip}$	$\overline{X} = \frac{1}{npq}\sum_{i=1}^{q}\sum_{j=1}^{p} n\overline{x}_{ij}$

- $\overline{x}_{ij}$: mean of each sample having level B_i and level A_j.
- $\overline{x}_{i.}$: mean of row i.
- $\overline{x}_{.j}$: mean of column j.
- $\overline{X}$: grand mean.

Variance table

$S^2_{ij} = \frac{n}{n-1}S^2_{e_{ij}}$	A_1	A_2	....	A_p
B_1	S^2_{11}	S^2_{12}		S^2_{1p}
B_2	S^2_{21}	S^2_{22}		S^2_{2p}
. . .				
B_q	S^2_{q1}	S^2_{q2}		S^2_{qp}

Estimated variance of the sample having level B_i and level A_j. The residual variance S_R^2 is the average of the estimated variances:

$$S_R^2 = \frac{1}{(n-1)pq} \sum_{\substack{1 \le i \le q \\ 1 \le j \le p}} (n-1) S_{ij}^2.$$

The total estimated variance S_T^2 is the estimated variance of all samples.

Decomposition of the mean

$$x_{ijk} - \overline{X} = \left(\overline{x}_{i.} - \overline{X}\right) + \left(\overline{x}_{.j} - \overline{X}\right) + \left(\overline{x}_{ij} - \overline{x}_{i.} - \overline{x}_{.j} + \overline{X}\right) + \left(x_{ijk} - \overline{x}_{ij}\right)$$

- $\left(\overline{x}_{i.} - \overline{X}\right)$: effect of factor B.
- $\left(\overline{x}_{.j} - \overline{X}\right)$: effect of factor A.
- $\left(\overline{x}_{ij} - \overline{x}_{i.} - \overline{x}_{.j} + \overline{X}\right)$: interaction effect.
- $\left(x_{ijk} - \overline{x}_{ij}\right)$: residual error.

From this we extract:

$$SS_T = SS_A + SS_B + SS_{AB} + SS_R.$$

- SS_A: sum of squares of deviations of factor-A group means from the grand mean.
- SS_B: sum of squares of deviations of factor-B group means from the grand mean.
- SS_{AB}: sum of squares due to interaction.
- SS_F: sum of squares due to factors.
- SS_R: sum of squares of residual deviations.

Source (Sum of Squares)	Formula	Degrees of Freedom	Mean Squares
SS_A	$\sum_{j=1}^{p} nq(\overline{x}_{.j} - \overline{X})^2$	$p-1$	$MS_A = \frac{SS_A}{p-1}$
SS_B	$\sum_{i=1}^{q} np(\overline{x}_{i.} - \overline{X})^2$	$q-1$	$MS_B = \frac{SS_B}{q-1}$
SS_{AB}	$\sum_{\substack{1\leq i\leq q \\ 1\leq j\leq p}} n(\overline{x}_{ij} - \overline{x}_{i.} - \overline{x}_{.j} + \overline{X})^2$	$(q-1)(p-1)$	$MS_{AB} = \frac{SS_{AB}}{(q-1)(p-1)}$
SS_F	$SS_A + SS_B + SS_{AB}$		Computed using the theorem
SS_R	$\sum_{\substack{1\leq i\leq q \\ 1\leq j\leq p}} (n-1)S_{ij}^2$	$(n-1)pq$	$MS_R = \frac{SS_R}{(n-1)pq}$
SS_T	$\sum_{\substack{1\leq i\leq q \\ 1\leq j\leq p}} npq(x_{ijk} - \overline{X})^2$	$npq-1$	$MS_T = \frac{SS_T}{npq-1}$

Theorem 1 (Analysis of Variance Theorem).

$$(npq-1)MS_T = (n-1)pq\,MS_R + (pq-1)MS_F.$$

Theorem 2 (Factorial Variance Decomposition Theorem).

$$(pq-1)MS_F = (p-1)MS_A + (q-1)MS_B + (q-1)(p-1)MS_{AB}.$$

Decision

Test	Decision statistic	Distribution	df.
H_{0A}	$F_A = \frac{MS_A}{MS_R} = \frac{SS_A/(p-1)}{SS_R/((n-1)pq)}$	Fisher–Snedecor	$(p-1;\,(n-1)pq)$
H_{0B}	$F_B = \frac{MS_B}{MS_R} = \frac{SS_B/(q-1)}{SS_R/((n-1)pq)}$	Fisher–Snedecor	$(q-1;\,(n-1)pq)$
H_{0AB}	$F_{AB} = \frac{MS_{AB}}{MS_R} = \frac{SS_{AB}/((q-1)(p-1))}{SS_R/((n-1)pq)}$	Fisher–Snedecor	$((q-1)(p-1);\,(n-1)pq)$

For the 3 hypotheses, we compute each time $F_{(\alpha,\text{df.})}$. If the decision statistic $F_{.} > F_{(\alpha,\text{df.})}$, then H_0 is rejected.

Exercise 1. *We study the activity of an enzyme in subjects as a function of age and sex. The results are as follows:*

	less than 12 years		*greater than 12 years*	
Boy	4.9	2.9	2.1	2.2
	2.7	3.9	1.1	2.9
	4.6	3.3	5	3.5
	5.9	4.8	2.4	4.4
	4.1	3.5	2.1	3
	7.2	6.1	3.9	5.6
Girl	4.5	6.9	2.4	3.6
	4	5.4	4.8	3.9
	1.9	3.6	5.5	5
	4.8	3.3	6.8	2.2
	7.5	5.8	3.1	5
	4.4	6	4.1	4.7

Does the mean enzymatic activity depend on age and sex?

Source	**Sum of Squares**	**df**	**F**	**Sig.**
Sex	6.092	1	3.077	0.086
Age	10.735	1	5.422	0.025
Sex × Age	1.577	1	0.796	0.377
Error	87.124	44		
Total	105.528	47		

Table 1: Two-way ANOVA for enzymatic activity (SPSS output)

Solution 1. *Factor A (Age): $p=2$ levels; Factor B (Sex): $q=2$ levels. Each cell contains $n=12$ observations, hence*

$$df_R = (n-1)pq = (12-1)\cdot 2\cdot 2 = 44.$$

Sums of squares.

$$SS_A = 10.735208, \quad SS_B = 6.091875, \quad SS_{AB} = 1.576875, \quad SS_R = 87.124167.$$

Mean squares and Fisher statistics..

$$MS_R = \frac{SS_R}{44} = \frac{87.124167}{44} = 1.980095.$$

$$MS_A = \frac{SS_A}{p-1} = \frac{10.735208}{1} = 10.735208, \qquad F_A = \frac{MS_A}{MS_R} = 5.421563.$$

$$MS_B = \frac{SS_B}{q-1} = \frac{6.091875}{1} = 6.091875, \qquad F_B = \frac{MS_B}{MS_R} = 3.076557.$$

$$MS_{AB} = \frac{SS_{AB}}{(p-1)(q-1)} = \frac{1.576875}{1} = 1.576875, \qquad F_{AB} = \frac{MS_{AB}}{MS_R} = 0.796363.$$

Decision. At the 5% level, the critical value is

$$F_{0.05}(1, 44) \approx 4.062.$$

- ***Sex:*** $F_{Sex} = 3.077 < 4.062 \Rightarrow$ *fail to reject* H_0*. The mean enzymatic activity does not depend significantly on sex.*
- ***Age:*** $F_{Age} = 5.422 > 4.062 \Rightarrow$ *reject* H_0*. The mean enzymatic activity depends significantly on age.*
- ***Interaction (Sex×Age):*** $F_{Int} = 0.796 < 4.062 \Rightarrow$ *fail to reject* H_0*. No significant interaction between age and sex.*

2. Two-way ANOVA without Replication

When only one observation is available for each factor combination, two-factor ANOVA is limited to the study of main effects, assuming implicitly that there is no interaction between factors. The variance decomposition follows the same principle as in the case with replications.

Source (Sum of Squares)	Formula	Degrees of Freedom	Mean Squares
SS_A	$q\sum_{j=1}^{p}(\overline{x}_{.j} - \overline{X})^2$	$p-1$	$MS_A = \frac{SS_A}{p-1}$
SS_B	$p\sum_{i=1}^{q}(\overline{x}_{i.} - \overline{X})^2$	$q-1$	$MS_B = \frac{SS_B}{q-1}$
SS_R	$\sum_{\substack{1\leq i\leq q \\ 1\leq j\leq p}} (x_{ij} - \overline{x}_{i.} - \overline{x}_{.j} + \overline{X})^2$	$(q-1)(p-1)$	$MS_R = \frac{SS_R}{(q-1)(p-1)}$
SS_T	$\sum_{\substack{1\leq i\leq q \\ 1\leq j\leq p}} (x_{ij} - \overline{X})^2$	$pq-1$	

Decision

Test	Decision statistic	Distribution	df.
H_{0A}	$F_A = \frac{MS_A}{CM_R} = \frac{SS_A/(p-1)}{SS_R/((q-1)(p-1))}$	Fisher–Snedecor	$(p-1;\ (q-1)(p-1))$
H_{0B}	$F_B = \frac{MS_B}{MS_R} = \frac{SS_B/(q-1)}{SS_R/((q-1)(p-1))}$	Fisher–Snedecor	$(q-1;\ (q-1)(p-1))$

For the 2 hypotheses, we compute each time $F_{(\alpha,\text{df.})}$. If $F_{.} > F_{(\alpha,\text{df.})}$, then H_0 is rejected.

Exercise 2. *A researcher studied the weight gain of* 3 *different animal breeds (for each breed, he took only one animal) as a function of* 3 *different doses of vitamin* B_{12} $(5, 10$ *and* $15\,\mu g/cm^2)$*. The results are:*

	R_1	R_2	R_3
D_1	1.26	1.21	1.19
D_2	1.29	1.23	1.23
D_3	1.38	1.27	1.22

Is the weight gain significantly different at the 5% *level:*

- *according to breed?*
- *according to dose?*

Solution 2. *The experiment involves two factors:*

- ***Breed*** *(3 levels:* R_1, R_2, R_3*),*
- ***Dose of vitamin*** B_{12} *(3 levels:* D_1, D_2, D_3*).*

Since only one observation is available for each combination of breed and dose, this is a ***two-way ANOVA without replication****. Therefore, only the main effects (breed and dose) can be tested.*

Step 1: Means
Row (dose) means:

$$\overline{x}_{1.} = 1.22, \quad \overline{x}_{2.} = 1.25, \quad \overline{x}_{3.} = 1.29.$$

Column (breed) means:

$$\overline{x}_{.1} = 1.31, \quad \overline{x}_{.2} = 1.24, \quad \overline{x}_{.3} = 1.21.$$

Overall mean:

$$\overline{X} = \frac{1}{9}\sum_{i=1}^{3}\sum_{j=1}^{3} x_{ij} = \frac{11.29}{9} = 1.254.$$

Step 2: Sums of Squares
Total sum of squares:

$$SS_T = \sum_{i=1}^{3}\sum_{j=1}^{3}(x_{ij} - \overline{X})^2 = 0.0266.$$

Dose effect:

$$SS_D = 3\sum_{i=1}^{3}(\overline{x}_{i.} - \overline{X})^2 = 0.00738.$$

Breed effect:

$$SS_B = 3\sum_{j=1}^{3}(\overline{x}_{.j} - \overline{X})^2 = 0.0109.$$

Error sum of squares:

$$SS_R = SS_T - SS_D - SS_B = 0.00832.$$

Step 3: Degrees of Freedom

$$\begin{array}{lcl} \textit{Dose} & : & 2, \\ \textit{Breed} & : & 2, \\ \textit{Residual} & : & 4, \\ \textit{Total} & : & 8. \end{array}$$

Step 4: Mean Squares and F-statistics

$$MS_D = \frac{SS_D}{2} = 0.00369, \qquad MS_B = \frac{SS_B}{2} = 0.00545,$$
$$MS_R = \frac{SS_R}{4} = 0.00208.$$

$$F_D = \frac{MS_D}{MS_R} = 1.77, \qquad F_B = \frac{MS_B}{MS_R} = 2.62.$$

Step 5: Statistical Decision
At the 5% significance level, the critical value is

$$F_{0.05}(2, 4) = 6.94.$$

Since

$$F_D < F_{0.05}(2, 4) \quad \textit{and} \quad F_R < F_{0.05}(2, 4),$$

neither effect is statistically significant.

Conclusion
At the 5% significance level, the two-way ANOVA without replication shows that:

- *weight gain does not differ significantly according to breed;*
- *weight gain does not differ significantly according to dose.*