Generalized Linear Models, Analysis of Variance

(Time Series, and Econometrics, not an exam topic in 2021)

Analysis of Variance (ANOVA)

ANOVA investigates special linear models, used for planning experiments or quality control. Here the matrix of the deterministic predictors is a so-called *design-matrix* with 0-1 entries indicating that which predictors influence the response at all. For testing hypotheses, we will intensively use the following theorem and its corollaries.

Theorem 1 (Fisher–Cochran) *Let $\mathbf{X} = (X_1, \dots, X_n)^T \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$ be random vector and the quadratic forms $Q = \mathbf{X}^T \mathbf{I}_n \mathbf{X} = \mathbf{X}^T \mathbf{X} = \sum_{i=1}^n X_i^2$ and $Q_j = \mathbf{X}^T \mathbf{A}_j \mathbf{X}$ $(j = 1, \dots, k)$ be such that they satisfy*

$$Q = Q_1 + Q_2 + \cdots + Q_k,$$

where the $n \times n$ symmetric matrix has rank n_j $(j = 1, \dots, k \le n)$. Then the random variables $Q_1, Q_2, \dots, Q_k$ are independent $\chi^2(n_1)$-, $\chi^2(n_2)$- ,..., $\chi^2(n_k)$-distributed if and only if

$$\sum_{j=1}^{k} n_j = n.$$

We also list some propositions which follow from the Fisher–Cochran theorem:

Proposition 1 *Consider $\mathbf{X} \in \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$. Then*

- *(a) With some symmetric matrix $\mathbf{A}$, $\mathbf{X}^T \mathbf{A} \mathbf{X}$ is χ^2-distributed if $\mathbf{A}^2 = \mathbf{A}$. Then the degree of freedom of the χ^2-distribution*$=\mathrm{rank}(\mathbf{A}) = \mathrm{tr}(\mathbf{A})$.
- *(b) With some symmetric matrices $\mathbf{A}$, $\mathbf{B}$, the random variables $\mathbf{X}^T \mathbf{A} \mathbf{X} \sim \chi^2(a)$ and $\mathbf{X}^T \mathbf{B} \mathbf{X} \sim \chi^2(b)$ are independent if $\mathbf{A}\mathbf{B} = \mathbf{0}$ (then obviously, $a + b \le n$).*
- *(c) If $Q = Q_1 + Q_2$, $Q \sim \chi^2(a)$, $Q_1 \sim \chi^2(b)$, and Q_2 is positive semidefinite, then $Q_2 \sim \chi^2(a - b)$; further, Q_1 and Q_2 are independent.*

ANOVA

In the one-way case, we investigate whether different treatments (conditions) influence significantly some continuous measurements (response). For example, whether the GDP differs significantly in different countries. If the continuous measurement is normally distributed, this is the generalization of the t-test for more than two groups.

In the two-way case, two kinds of treatments are given on different levels, and we have a two-way classified table of continuous measurements. For example, whether the GDP differs significantly in different countries and under different economic regulations. Then we may investigate the effect of the countries, the effect of the regulations, and the interaction between them. We discuss these models with precise formulas.

ANOVA models:

- *One-way ANOVA*: Our sample is taken in k different groups: X_{ij}, $j = 1, \dots, n_i$, $i = 1, \dots, k$, and the sample size is $n = \sum_{i=1}^k n_i$. Assume that in group i our observations follow $\mathcal{N}(b_i, \sigma^2)$-distribution. It is important that the observations are independent with the same variance (homoskedasticity). We want to test the null-hypothesis

  $$H_0 : b_1 = b_2 = \cdots = b_k,$$

  which is the generalization of the t-test for k groups. We use the decomposition $b_i = \mu + a_i$, where

  $$\mu = \frac{1}{n} \sum_{i=1}^k n_i b_i, \qquad a_i = b_i - \mu \quad (i = 1, \dots, k).$$

  Obviously,

  $$\sum_{i=1}^k n_i a_i = 0.$$

  With this notation, our model is

  $$X_{ij} = \mu + a_i + \varepsilon_{ij} \qquad (j = 1, \dots, n_i;\ i = 1, \dots, k) \tag{1}$$

  where $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ are i.i.d. error terms. This model is, in fact, a linear model with the n-dimensional vectors

  $$\mathbf{Y} := (X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}, \dots, X_{k1}, \dots, X_{kn_k})^T$$
  $$\underline{\varepsilon} := (\varepsilon_{11}, \dots, \varepsilon_{1n_1}, \varepsilon_{21}, \dots, \varepsilon_{2n_2}, \dots, \varepsilon_{k1}, \dots, \varepsilon_{kn_k})^T$$

  and parameter-vector $\mathbf{a} = (a_1, \dots, a_k)^T$. With these, the above model is

  $$\mathbf{Y} = \mathbf{B}\mathbf{a} + \mu \mathbf{1} + \underline{\varepsilon}$$

  where the vector $\mathbf{1} \in \mathbb{R}^n$ has all 1 coordinates, and the $n \times k$ design matrix $\mathbf{B}$ is such that in its i-th column it contains all 0's, except the block i, where it contains 1's. This ensures that in group i only the parameter a_i appears in the model equation (1).

The parameters are estimated by the method of least squares: we minimize

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\varepsilon_{ij}^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\mu-a_i)^2.$$

The least square estimates of the parameters are

$$\hat{\mu}=\bar{X}_{..} \quad \text{and} \quad \hat{a}_i=\bar{X}_{i.}-\bar{X}_{..} \quad (i=1,\dots,k),$$

where

$$\bar{X}_{i.}=\frac{1}{n_i}\sum_{j=1}^{n_i}X_{ij} \quad (i=1,\dots,k) \quad \text{and} \quad \bar{X}_{..}=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}X_{ij}.$$

The minimum is

$$SSE=Q_e=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\hat{m}-\hat{a}_i)^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i.})^2.$$

With

$$SSR=Q_a=\|\mathbf{B}\hat{\mathbf{a}}\|^2=\sum_{i=1}^{k}n_i\hat{a}_i^2=\sum_{i=1}^{k}n_i(\bar{X}_{i.}-\bar{X}_{..})^2,$$

we can decompose the total variation of the sample ($SST=Q$) into between-groups ($SSR=Q_a$) and within-groups ($SSE=Q_e$) variation as follows:

$$\begin{aligned}
Q&=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{..})^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}[(X_{ij}-\bar{X}_{i.})+(\bar{X}_{i.}-\bar{X}_{..})]^2=\\
&=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i.})^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{X}_{i.}-\bar{X}_{..})^2=\\
&=\sum_{i=1}^{k}n_i(\bar{X}_{i.}-\bar{X}_{..})^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i.})^2=Q_a+Q_e.
\end{aligned}$$

This decomposition is summarized in the 1-way ANOVA table:

Cause of the dispersion	Sum of squares	Degrees of freedom	Empirical variance
Between groups	$Q_a=\sum_{i=1}^{k}n_i(\bar{X}_{i.}-\bar{X}_{..})^2$	$k-1$	$s_a^2=\frac{Q_a}{k-1}$
Within groups	$Q_e=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i.})^2$	$n-k$	$s_e^2=\frac{Q_e}{n-k}$
Total	$Q=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{..})^2$	$n-1$	-

In the above model, we first investigate the null-hypothesis $\mu=0$. If we reject it, we investigate the null-hypothesis, that there is no difference between the groups:

$$H_0 \,:\, a_1=\cdots=a_k=0, \quad \text{briefly,} \quad \mathbf{a}=\mathbf{0}.$$

In view of the Fisher–Cochran theorem and its consequences, $Q_e \sim \sigma^2\chi^2(n-k)$, irrespective whether H_0 holds or not. However, the expectation of the linear expressions in Q_a is

$$\mathbb{E}(\bar{X}_{i.} - \bar{X}_{..}) = \mathbb{E}(\bar{X}_{i.}) - \mathbb{E}(\bar{X}_{..}) = a_i - \frac{1}{n}\sum_{j=1}^{k} n_j a_j \qquad (i = 1, \ldots, k)\,,$$

which can be zero for all i only if H_0 holds. In this case, with the Fisher–Cochran theorem and its consequences, $Q_a \sim \sigma^2\chi^2(k-1)$; further, Q_e and Q_a are independent of each other. Observe that the degrees of freedom in the decomposition

$$Q = Q_a + Q_e$$

are added together:

$$n - 1 = (k-1) + (n-k).$$

Therefore, with the notation

$$s_a^2 = \frac{Q_a}{k-1} \qquad \text{and} \qquad s_e^2 = \frac{Q_e}{n-k},$$

the test statistic

$$F = \frac{s_a^2}{s_e^2} = \frac{Q_a}{Q_e} \cdot \frac{n-k}{k-1} \sim \mathcal{F}(k-1, n-k)$$

follows Fisher F-distribution with the above degrees of freedom under H_0. Summarizing, if $F \geq F_\alpha(k-1, n-k)$, i.e., the between-group variances are significantly larger than the within-group-ones, then we reject H_0 with significance α.

- *Bartlett-test* for testing equality of the variances of the groups:

  $$H_0 \,:\, \sigma_1 = \cdots = \sigma_k.$$

  Based on the above grouped sample, the test statistic is

  $$B^2 = \frac{2.3026}{c}\left(f \lg S^{*2} - \sum_{i=1}^{k} f_i \lg S_i^{*2}\right),$$

  where $f_i = n_i - 1$ $(i = 1, \ldots, k)$; $f = \sum_{i=1}^{k} f_i$; S_1^{*2}, $\ldots$,S_k^{*2} are the corrected empirical variance within the groups, and

  $$S^{*2} = \frac{1}{f}\sum_{i=1}^{k} f_i S_i^{*2}, \quad c = 1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{f_i} - \frac{1}{f}\right).$$

  Bartlett proved that for 'large' sample sizes, B^2 asymptotically follows $\chi^2(k-1)$-distribution. Therefore, if $B^2 \geq \chi^2_\alpha(k-1)$, then we reject H_0 with significance α.

- *Two-way ANOVA without interaction*: We have two-way classified data in $k \cdot p$ groups. Here one observation per group suffices. Let X_{ij} denote the continuous measurements ($i = 1, \ldots, k$; $j = 1, \ldots, p$), the sample size is $n = kp$. We assume that there is no interaction between the two treatments (based on the levels of which we form the groups).

  In our model, the independent, homoskedastic sample entries have $X_{ij} \sim \mathcal{N}(\mu+a_i+b_j, \sigma^2)$ distribution. Therefore, our linear model is the following:

  $$X_{ij} = \mu + a_i + b_j + \varepsilon_{ij}, \qquad (i = 1, \ldots, k;\ j = 1, \ldots, p) \tag{2}$$

  where $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ are i.i.d. errors. The parameters a_i's and b_j's denote the non-interacting effects of the levels of the two treatments. We can assume (with the choice of μ) that

  $$\sum_{i=1}^{k} a_i = 0 \qquad \text{and} \qquad \sum_{j=1}^{p} b_j = 0.$$

  Here we do not specify the design matrices, but (2) also fits into the framework of linear models.

  Minimizing the objective function

  $$\sum_{i=1}^{k} \sum_{j=1}^{p} \varepsilon_{ij}^2 = \sum_{i=1}^{k} \sum_{j=1}^{p} (X_{ij} - \mu - a_i - b_j)^2$$

  by the method of least squares, we obtain the following estimates of the parameters:

  $$\begin{aligned} \hat{\mu} &= \bar{X}_{..}\,, \\ \hat{a}_i &= \bar{X}_{i.} - \bar{X}_{..} \qquad (i = 1, \ldots, k)\,,\,, \\ \hat{b}_j &= \bar{X}_{.j} - \bar{X}_{..} \qquad (j = 1, \ldots, p)\,, \end{aligned}$$

  where

  $$\begin{aligned} \bar{X}_{i.} &= \frac{1}{p} \sum_{j=1}^{p} X_{ij} \quad (i = 1, \ldots, k) \\ \bar{X}_{.j} &= \frac{1}{k} \sum_{i=1}^{k} X_{ij} \quad (j = 1, \ldots, p) \\ \bar{X}_{..} &= \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{p} X_{ij}. \end{aligned}$$

  With this, the minimum of our objective function is

  $$SSE = Q_e = \sum_{i=1}^{k} \sum_{j=1}^{p} (X_{ij} - \hat{\mu} - \hat{a}_i - \hat{b}_j)^2.$$

  Further, we have the decomposition

  $$Q = Q_a + Q_b + Q_e \tag{3}$$

of the total variance ($SST = Q$) into the variances caused by the a-effects, b-effects and the error (Q_a, Q_b, Q_e, where $SSR = Q_a + Q_b$).

This decomposition is summarized in the 2-way ANOVA table (without interaction):

Cause of the dispersion	Sum of squares	Degree of freedom	Empirical variance
a-effects	$Q_a = p\sum_{i=1}^{k}(\bar{X}_{i.} - \bar{X}_{..})^2$	$k-1$	$s_a^2 = \frac{Q_a}{k-1}$
b-effects	$Q_b = k\sum_{j=1}^{p}(\bar{X}_{.j} - \bar{X}_{..})^2$	$p-1$	$s_b^2 = \frac{Q_b}{p-1}$
Random error	$Q_e = \sum_{i=1}^{k}\sum_{j=1}^{p}(X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2$	$(k-1)(p-1)$	$s_e^2 = \frac{Q_e}{(k-1)(p-1)}$
Total	$Q = \sum_{i=1}^{k}\sum_{j=1}^{p}(X_{ij} - \bar{X}_{..})^2$	$kp-1$	-

If we have rejected the null-hypothesis $\mu = 0$, we compare the levels of both treatments, separately. To compare a-effects, we investigate

$$H_{0a} :\ a_1 = a_2 = \cdots = a_k = 0, \quad \text{briefly,} \quad \mathbf{a} = \mathbf{0}.$$

To compare b-effects, we investigate

$$H_{0b} :\ b_1 = b_2 = \cdots = b_p = 0, \quad \text{briefly,} \quad \mathbf{b} = \mathbf{0}.$$

In view of the Fisher–Cochran theorem and its consequences, since the degrees of freedoms of the terms in (3) are added together,

$$kp - 1 = (k-1) + (p-1) + (k-1)(p-1),$$

we have the following facts:

- $Q_e/\sigma^2 \sim \chi^2((k-1)(p-1))$, irrespective whether the above hypotheses hold or not.
- Under H_{0a}, $Q_a/\sigma^2 \sim \chi^2(k-1)$ and is independent of Q_e.
- Under H_{0b}, $Q_b/\sigma^2 \sim \chi^2(p-1)$ and is independent of Q_e.

Therefore, under H_{0a}, the test statistic

$$F_a = \frac{s_a^2}{s_e^2} \sim \mathcal{F}(k-1, (k-1)(p-1)),$$

and if $F_a \geq F_\alpha(k-1, (k-1)(p-1))$, then we reject H_{0a} with significance α.

Likewise, Therefore, under H_{0b}, the test statistic

$$F_b = \frac{s_b^2}{s_e^2} \sim \mathcal{F}(p-1, (k-1)(p-1)),$$

and if $F_b \geq F_\alpha(p-1, (k-1)(p-1))$, then we reject H_{0b} with significance α.

- *Two-way ANOVA with interaction*: Here we also have two-way classified data in $k \cdot p$ groups, but we have more than one (say, n) observations per cell, since there is interaction between the two treatments. The independent, homoskedastic sample entries are X_{ijl} $(i = 1, \dots, k;\ j = 1, \dots, p;\ l = 1, \dots, n)$. The sample size is kpn. Supposing that $X_{ijl} \sim \mathcal{N}(\mu + a_i + b_j + c_{ij}, \sigma^2)$, our linear model is the following:

$$X_{ijl} = \mu + a_i + b_j + c_{ij} + \varepsilon_{ijl}, \qquad (i = 1, \dots, k;\ j = 1, \dots, p),$$

where $\varepsilon_{ijl} \sim \mathcal{N}(0, \sigma^2)$ are i.i.d. errors, a_i's and b_j's denote the effects of the two treatments, whereas c_{ij}'s are the interactions. We assume that

$$\sum_{i=1}^{k} a_i = 0, \qquad \sum_{j=1}^{p} b_j = 0,$$
$$\sum_{i=1}^{k} c_{ij} = 0 \qquad (j = 1, \dots, p) \quad \text{és}$$
$$\sum_{j=1}^{p} c_{ij} = 0 \qquad (i = 1, \dots, k).$$

This model is also a linear one.

By the method of least squares, the minimum of

$$\sum_{i=1}^{k} \sum_{j=1}^{p} \sum_{l=1}^{n} \varepsilon_{ijl}^2 = \sum_{i=1}^{k} \sum_{j=1}^{p} \sum_{l=1}^{n} (X_{ijl} - \mu - a_i - b_j - c_{ij})^2$$

with respect to the parameters $\mu, a_1, \dots, a_k, b_1, \dots, b_p$, under the above constraints, is attained at

$$\begin{aligned}
\hat{\mu} &= \bar{X}_{...}\,, \\
\hat{a}_i &= \bar{X}_{i..} - \bar{X}_{...} \qquad (i = 1, \dots, k)\,, \\
\hat{b}_j &= \bar{X}_{.j.} - \bar{X}_{...} \qquad (j = 1, \dots, p)\,, \\
\hat{c}_{ij} &= \bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...} \qquad (i = 1, \dots, k; \quad j = 1, \dots, p)\,,
\end{aligned}$$

where

$$\bar{X}_{i..} = \frac{1}{pn} \sum_{j=1}^{p} \sum_{l=1}^{n} X_{ijl} \quad (i = 1, \dots, k)$$

$$\bar{X}_{.j.} = \frac{1}{kn} \sum_{i=1}^{k} \sum_{l=1}^{n} X_{ijl} \quad (j = 1, \dots, p)$$

$$\bar{X}_{ij.} = \frac{1}{n} \sum_{l=1}^{n} X_{ijl} \qquad (i = 1, \dots, k;\ j = 1, \dots, p)$$

$$\bar{X}_{...} = \frac{1}{kpn} \sum_{i=1}^{k} \sum_{j=1}^{p} \sum_{l=1}^{n} X_{ijl}.$$

The least square estimates of the parameters are

$$\begin{aligned}
\hat{m} &= \bar{X}_{...}\,, \\
\hat{a}_i &= \bar{X}_{i..} - \bar{X}_{...} \qquad (i = 1, \dots, k)\,, \\
\hat{b}_j &= \bar{X}_{.j.} - \bar{X}_{...} \qquad (j = 1, \dots, p)\,, \\
\hat{c}_{ij} &= \bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...} \qquad (i = 1, \dots, k; \quad j = 1, \dots, p)\,,
\end{aligned}$$

and the minimum is

$$SSE = Q_e = \sum_{i=1}^{k} \sum_{j=1}^{p} \sum_{l=1}^{n} (X_{ijl} - \hat{m} - \hat{a}_i - \hat{b}_j - \hat{c}_{ij})^2.$$

For the decomposition

$$Q = Q_a + Q_b + Q_c + Q_e$$

we again have the ANOVA-table:

Cause of dispersion	Sum of squares	Degree of freedom	Empirical variance
a-effects	$Q_a = pn \sum_{i=1}^{k} (\bar{X}_{i..} - \bar{X}_{...})^2$	$k-1$	$s_a^2 = \frac{Q_a}{k-1}$
b-effects	$Q_b = kn \sum_{j=1}^{p} (\bar{X}_{.j.} - \bar{X}_{...})^2$	$p-1$	$s_b^2 = \frac{Q_b}{p-1}$
ab-interaction	$Q_c = n \sum_{i=1}^{k} \sum_{j=1}^{p} (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...})^2$	$(k-1)(p-1)$	$s_c^2 = \frac{Q_c}{(k-1)(p-1)}$
Random error	$Q_e = \sum_{i=1}^{k} \sum_{j=1}^{p} \sum_{l=1}^{n} (X_{ijl} - \bar{X}_{ij.})^2$	$kp(n-1)$	$s_e^2 = \frac{Q_e}{kp(n-1)}$
Total	$Q = \sum_{i=1}^{k} \sum_{j=1}^{p} \sum_{l=1}^{n} (X_{ijl} - \bar{X}_{...})^2$	$kpn-1$	-

After rejecting the null-hypothesis $\mu = 0$, we investigate the interaction:

$$H_{0ab} : \, c_{ij} = 0, \quad (i = 1, \dots, k; \, j = 1, \dots, p).$$

If we accept it (no interaction), we investigate separately

$$H_{0a} : \, a_1 = a_2 = \cdots = a_k = 0$$

and

$$H_{0b} : \, b_1 = b_2 = \cdots = b_p = 0.$$

Using the Fisher–Cochran theorem and its consequences, further, the additivity of the degrees of freedoms,

$$kpn - 1 = (k-1) + (p-1) + (k-1)(p-1) + kp(n-1),$$

we have the following facts:

– $Q_e/\sigma^2 \sim \chi^2(kp(n-1))$, always.

– Under H_{0a}, $Q_a/\sigma^2 \sim \chi^2(k-1)$ and is independent of Q_e.
– Under H_{0b}, $Q_b/\sigma^2 \sim \chi^2(p-1)$ and is independent of Q_e.
– Under H_{0ab}, $Q_c/\sigma^2 \sim \chi^2((k-1)(p-1))$ and is independent of Q_e.

Therefore, we have the following test statistics: Under H_{0ab},

$$F_{ab} = \frac{s_c^2}{s_e^2} \sim \mathcal{F}((k-1)(p-1), kp(n-1)),$$

Under H_{0a},

$$F_a = \frac{s_a^2}{\tilde{s}_e^2} \sim \mathcal{F}(k-1, kpn-k-p+1).$$

Under H_{0b},

$$F_b = \frac{s_b^2}{\tilde{s}_e^2} \sim \mathcal{F}(k-1, kpn-k-p+1).$$

Then you can make the conclusions at significance α.

Note that the are so-called mixed ANOVA models, with different number of observations per cell, or with more than two factors. In these cases, we have to build up a design-matrix and use the Gauss normal equations to estimate the parameters or organize the experiments is simple patterns, e.g. Latin squares in the 3-way case.

- The ANOVA model can be extended to multivariate, grouped observations. In the 1-way Multivariate Analysis of Variance (MANOVA) setup, our p-variate measurements

$$\mathbf{Y}_{ij} \sim \mathcal{N}_p(\mathbf{m} + \mathbf{a}_i, \mathbf{C}) \qquad (j = 1, \ldots, n_i;\ i = 1, \ldots, k)$$

are assigned to k different groups, where $\sum_{i=1}^k \mathbf{a}_i = \mathbf{0}$ is assumed. Our inference is based on the decomposition

$$\mathbf{T} = \mathbf{B} + \mathbf{W}$$

of n times the $p \times p$ sample covariance matrix into *between- and within-group covariance matrices* in the following way:

$$\begin{aligned} \mathbf{T} &= \sum_{i=1}^k \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{..})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{..})^T \\ \mathbf{B} &= \sum_{i=1}^k n_i (\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..})(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..})^T \\ \mathbf{W} &= \sum_{i=1}^k \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})^T \end{aligned} \tag{4}$$

where $\bar{\mathbf{Y}}_{..} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} \mathbf{Y}_{ij}$ is the sample mean vector, while $\bar{\mathbf{Y}}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{Y}_{ij}$ is the mean vector of group i $(i = 1, \ldots, k)$.