

AN ELEMENTARY PROOF OF FISHER-COCHRAN THEOREM USING A GEOMETRICAL APPROACH

Lucas Monteiro CHAVES¹
Devanil Jaques de SOUZA²

- **ABSTRACT:** The classical Fisher-Cochran theorem is a fundamental result in many areas of statistics as analysis of variance and hypothesis tests. In general this theorem is proved with linear algebraic arguments. An elementary proof is present, based strongly on geometrical concepts as linear subspaces and orthogonal projections, which may improve our intuition about the result.
- **KEYWORDS:** Chi-squared distribution; quadratic forms; orthogonal projections.

1 Introduction

Sum of squares are very important in many branches of statistics as in least-squares methods, analysis of variance, regression analysis, experimental design, etc. It is usual in experimental context that many factors add up to cause random error effects and it is desirable to quantify the contribution of each factor. Thus, the general idea is to split the sum of squares into a number of particular sums of squares, each corresponding to a different cause of variability. This procedure has its probabilistic behavior described by a classical result, the Fisher-Cochran theorem, whose demonstration, strongly based on linear algebra, may be found in textbooks, as in Rao (2008), in a general situation or a demonstration sketch in Wichura (2006) and Rencher and Shaalje (2008). This note intends to use the maximum of geometry and a minimum of linear algebra to obtain a simple alternative proof. Geometry always give us some intuition about the result and so we hope to be useful for the students.

¹Universidade Federal de Lavras - UFLA, Departamento de Ciências Exatas, CEP: 37200-000, Lavras, MG, Brasil, E-mail: *lucas@ufla.br*

²Universidade Federal de Lavras - UFLA, Departamento de Estatística, CEP: 37200-000, Lavras, MG, Brasil, E-mail: *devaniljaques@ufla.br*

The paper is organized as follows: section 2 with some basic linear algebra results, section 3 with some basic probability theory results, section 4 with the geometrical proof and section 5 with conclusions.

2 Some basic linear algebra results

Let us introduce the notation and definitions used for the linear algebra results.

The projection of a linear space \mathbb{R}^n onto a subspace $V \subset \mathbb{R}^n$, $A : \mathbb{R}^n \rightarrow V \subset \mathbb{R}^n$, is an idempotent linear transformation, that is, for any vector $v \in \mathbb{R}^n$, $Av \in V$ and $A(Av) = A^2v = Av$. That is, any $v \in \mathbb{R}^n$ is projected in V by the transformation A and any vector in V is fixed by A .

Since $A(v - Av) = Av - A^2v = Av - Av = 0$, the kernel of the transformation may be expressed as $KerA = \{v - Av; v \in \mathbb{R}^n\}$ and the vector space \mathbb{R}^n by the direct sum $\mathbb{R}^n = V \oplus KerA$. If the subspace $KerA$ is perpendicular to V , the matrix A is said to be an *orthogonal projection*. The main result we need about these projectors is:

Proposition: A projection matrix $A : \mathbb{R}^n \rightarrow V \subset \mathbb{R}^n$ is orthogonal if, and only if, it is symmetric.

Proof:

Observe that for any pair (v, w) , $v, w \in \mathbb{R}^n$, $\langle Av, w \rangle = \langle v, A^T w \rangle$ and if we have

$$\langle Av, w \rangle = \langle v, Aw \rangle \forall v, w \quad \text{then} \quad A^T = A.$$

(\Rightarrow) Suppose that the projection is orthogonal. Then, $\forall v, w \in \mathbb{R}^n$:

$$(a) \quad 0 = \langle Av - v, Aw \rangle = \langle Av, Aw \rangle - \langle v, Aw \rangle \Rightarrow \langle Av, Aw \rangle = \langle v, Aw \rangle.$$

$$(b) \quad 0 = \langle Aw - w, Av \rangle = \langle Aw, Av \rangle - \langle w, Av \rangle \Rightarrow \langle Aw, Av \rangle = \langle w, Av \rangle.$$

Thus, (a) and (b) imply that $\langle v, Aw \rangle = \langle w, Av \rangle$ and therefore A is symmetric.

(\Leftarrow) Suppose now that matrix A is symmetric. Then,

$$\begin{aligned} \langle v - Av, Aw \rangle &= \langle v, Aw \rangle - \langle Av, Aw \rangle \\ &= \langle v, Aw \rangle - \langle v, A^T Aw \rangle \\ &= \langle v, Aw \rangle - \langle v, A^2 w \rangle \quad (A \text{ is symmetric}) \\ &= \langle v, Aw \rangle - \langle v, Aw \rangle = 0 \quad (A \text{ is a projection}) \end{aligned}$$

Thus $v - Av$ and Aw are orthogonal vectors, that is, $KerA \perp V$ and therefore A is orthogonal.

If A is a $n \times n$ matrix and y is a $n \times 1$ vector, then the product $y^T A y$ is known as a quadratic form. Observe that A can always be supposed to be symmetric

because, if it is not, it is possible to construct the same quadratic form with a symmetric matrix (would you like to verify?). If one calls a_i the i -th column of A , then the quadratic form may be rewritten as $y^T A y = \langle (y_1, \dots, y_n), (a_1^T y, \dots, a_n^T y) \rangle = \sum_{i=1}^n y_i a_i^T y$. For theory of quadratic form and Gauss-Markov estimator see Eaton (1970) and Haberman (1975).

Recall that the rank of a matrix is the dimension of its image subspace and is given by the number of its linearly independent columns.

3 Some basic results on probability theory

Let us introduce the notation and definitions used for the Probability theory results.

If X is a standard normal random variable, $X \sim \text{Normal}(0,1)$, and X_1, X_2, \dots, X_n is a sequence on n independent copies of X then $X_1^2 + \dots + X_n^2$ has chi-square density with n degrees of freedom, with notation $(X_1^2 + \dots + X_n^2) \sim \chi_n^2$.

If $y^T = (y_1, \dots, y_n)$ is a vector of independent random variables, with $y_j \sim \text{Normal}(0,1)$, $j = 1, \dots, n$, then $\|y\|^2 \sim \chi_n^2$.

Given a vector of constants $a^T = (a_1, \dots, a_n)$, the linear combination $a^T y = y^T a = a_1 y_1 + \dots + a_n y_n$ is normal, with zero mean and variance $\|a\|^2$.

If b is another $n \times 1$ vector of constants, then $\text{cov}(a^T y, b^T y) = a^T b$. Finally, if a and b are orthogonal, then $a^T y$ and $b^T y$ are independent (BAILEY, 2008).

4 The Fisher-Cochran theorem

This fundamental results is known since the 1920 (FISHER, 1973), and one of it's versions is:

Theorem: If $y \sim \text{Normal}_n(0, I)$ and A_i ($i = 1, \dots, k$) are symmetric and idempotent matrices with $\text{rank}(A_i) = r_i$ and $\sum_{i=1}^k r_i = n$ then the following statements are equivalent:

- a) $A_i A_j = 0$ ($i \neq j$).
- b) The quadratic form $y^T A_i y$ ($i = 1, 2, \dots, k$) are independent random variables, distributed as chi-square with r_i degrees of freedom.

Proof:

a) \Rightarrow b)

The matrices A_i ($i = 1, \dots, k$) are projections, the conditions $A_i A_j = 0$ ($i \neq j$) imply that if V_i is the image of A_i then V_i and V_j orthogonal subspaces and as $\sum_{i=1}^k r_i = n$, \mathbb{R}^n is the direct sum $\mathbb{R}^n = A_1 \oplus A_2 \oplus \dots \oplus A_k$.

The basic idea for the proof is to consider a orthogonal basis $\{v_1, v_2, \dots, v_n\}$ adapted to this decomposition, that is, if $v_s \in V_j$ then or $v_{s+1} \in V_j$ or $v_{s+1} \in V_{j+1}$. Observe that A_i acts as identity over V_i , that is, if $v \in V_i$ $A_i v = v$. In this case, restrict to $v \in V_i$, $v^T A_i v = v^T v = \|v\|^2$. In relation to this adapted basis, any vector y may be decompose as $y^T v_1 + y^T v_2 + \dots + y^T v_n$. Therefore $A_i y$, the projection of vector y in V_i is given by $A_i y = \sum_{v_s \in V_i} (y^T v_s) v_s$. Such fact implies that

$$y = \sum_{i=1}^k A_i y \text{ and, therefore, we have the matrix decomposition } I = A_1 + A_2 + \dots + A_k.$$

As consequence:

$$\|y\|^2 = y^T I y = y^T (A_1 + A_2 + \dots + A_k) y = y^T A_1 y + y^T A_2 y + \dots + y^T A_k y.$$

Let us analyze the distribution of each quadratic form $y^T A_i y$. Using the fact that $A_i y = \sum_{v_s \in V_i} (y^T v_s) v_s$ we have $y^T A_i y = \sum_{v_s \in V_i} (y^T v_s)^2$. Now we can use the results in section 3:

$y \sim Normal_n(0, I)$ then $y^T v_s \sim Normal_1(0, 1)$, and for $s \neq m$, $y^T v_s$ and $y^T v_m$ are independent. Therefore $(y^T v_s)^2$ is chi-square distributed with one degree of freedom and $y^T A_i y = \sum_{v_s \in V_i} (y^T v_s)^2$ is chi-square distributed with $r_i = \dim(V_i)$ degree of freedom.

Thus indeed, condition **a)** implies condition **b)**.

b) \Rightarrow a)

In another words, we have to show that the subspaces V_i ($i = 1, 2, \dots, k$) are mutually orthogonal. Without loss of generality, we will proof that V_1 and V_2 are orthogonal. Take a generic vector $y \in V_2$ (Figure 1). By the orthogonal direct sum $\mathbb{R}^n = V_1 \oplus V_1^\perp$, we have the orthogonal decomposition $y = y_1 + y_1^\perp$. We have to proof that the component $y_1 = 0$.

Note that,

$$y^T A_2 y = y^T y = \|y\|^2 = \|y_1\|^2 + \|y_1^\perp\|^2 = y_1^T A_1 y_1 + \|y_1^\perp\|^2.$$

Taking the covariance between these quadratic forms independents

$$0 = \text{cov}(y^T A_2 y, y_1^T A_1 y_1) = \text{cov}(y_1^T A_1 y_1, y_1^T A_1 y_1) + \text{cov}(\|y_1^\perp\|^2, y_1^T A_1 y_1).$$

But y_1 and y_1^\perp are orthogonal normal random vectors. Therefore they are independent and so the quadratic forms $\|y_1^\perp\|^2$ and $y_1^T A_1 y_1$ are independent, that is, $\text{cov}(\|y_1^\perp\|^2, y_1^T A_1 y_1) = 0$ which implies that

$$\text{cov}(y_1^T A_1 y_1, y_1^T A_1 y_1) = \text{var}(y_1^T A_1 y_1) = \text{var}(\|y_1^\perp\|^2) = 0.$$

Therefore $y_1 = 0$ which implies that V_1 and V_2 are mutually orthogonal.

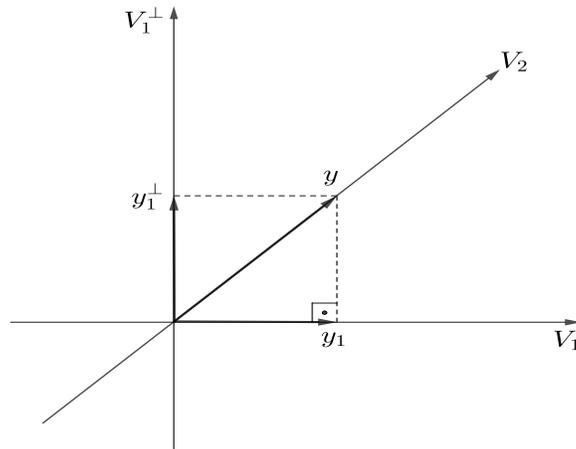


Figure 1 - Orthogonality between V_1 and V_2 .

5 Conclusion

It is possible, using orthogonal projection and subspaces, to give, avoiding intense linear algebra, a simple and intuitive demonstration for the classical Fisher-Cochran theorem. This approach may have some pedagogical value and certainly can be extended for understand applications and further generalizations related with sum of squares.

For a different approach to the Fisher-Cochran theorem we suggests the references below.

6 Acknowledgements

We would like to thank referees and editors for their valuable suggestions.

CHAVES, L. M.; SOUZA, D. J. Uma prova elementar do teorema de Fisher-Cochran usando uma abordagem geométrica. *Rev. Bras. Biom.*, Lavras, v.37, n.3, p.372-377, 2019.

- **RESUMO:** O teorema clássico de Fisher-Cochran é um resultado fundamental em muitas áreas da estatística tais como análise de variância e testes de hipóteses. Em geral, este teorema é provado com argumentos algébricos lineares. Neste artigo apresenta-se uma prova elementar baseada fortemente em conceitos geométricos tais como subespaços lineares e projeções ortogonais que podem melhorar nossa intuição sobre o resultado.
- **PALAVRAS-CHAVE:** Distribuição do qui-quadrado; formas quadráticas; projeções ortogonais.

References

- BAILEY, R. A. *Design of comparative experiments*. New York: Cambridge University Press, 2008.
- EATON, M. L. Gauss-Markov estimation for multivariate linear models: a coordinate-free approach. *Ann. Math. Statist.*, v.41, n.2, p.528-538, 1970.
- FISHER, L. An alternative approach to Cochran's theorem. *The American Statistician*, v.27, n.3, p.109-109, 1973.
- HABERMAN, S. J. How much do Gauss-Markov and least square differ? A coordinate-free approach. *Ann. Statist.*, v.3, n.4, p.982-990, 1975.
- RAO, R. *Linear statistical inference and its applications*. 2.ed. New York: John Wiley, 2008.
- RENCHEER, A. C.; SHAALJE, G. B. *Linear models in statistics*. New York: John Wiley, 2008.
- WICHURA, M. J. *The coordinate-free approach to linear models*. Cambridge: Cambridge University Press, 2006.

Received on 18.09.2018.

Approved after revised on 10.07.2019.